UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

Tomografia da Terra pela Emissão de Geo-Neutrinos

Leonardo Estêvão Schendes Tavares

Orientador: Prof. Dr. Nestor Felipe Caticha Alfonso (IF-USP)
Co-Orientador: Prof. Dr. Renata Zukanovich Funchal (IF-USP)

Dissertação de mestrado apresentada ao Instituto
de Física para a obtenção do título de Mestre em
Ciências.

Banca Examinadora:

Prof. Dr. Nestor Felipe Caticha Alfonso (Orientador, IF-USP)
Prof. Dr. Oscar José Pinto Éboli (If-USP)
Prof. Dr. Júlio Michael Stern (IME-USP)

São Paulo
2014

-
-
-

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Tavares, Leonardo Estevão Schendes

Tomography of the Earth by geo-neutrino emission \
Tomografia da terra pela emissão de geo-neutrinos. São Paulo,
2014.

Dissertação (Mestrado) – Universidade de São Paulo.
Instituto de Física. Depto. Física Geral

Orientador: Prof. Dr. Nestor Felipe Caticha Alfonso

Área de Concentração: Física

Unitermos: 1. Geo-neutrinos; 2. Bayes; 3. Inferência; 4. Terra;
5. Tomografia

USP/IF/SBI-061/2014

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

Tomography of the Earth by Geo-Neutrino Emission

Leonardo Estêvão Schendes Tavares

Advisor: Prof. Dr. Nestor Felipe Caticha Alfonso (IF-USP)
Co-advisor: Prof. Dr. Renata Zukanovich Funchal (IF-USP)

Dissertation submitted to the Institute of Physics of
the University of São Paulo for the obtainment of the
title of Master of Science.

Examining Committee:

Prof. Dr. Nestor Felipe Caticha Alfonso (Advisor, IF-USP)
Prof. Dr. Oscar José Pinto Éboli (If-USP)
Prof. Dr. Júlio Michael Stern (IME-USP)

São Paulo
2014

Abstract / Resumo

Geo-neutrinos are electronic anti-neutrinos originated from the beta decay process of some few elements in the decay chains of ^{232}Th and ^{238}U present in Earth's interior. Recent experimental measurements of these particles have been generating great expectations towards a new way for investigating directly the interior of the planet. It is a new multidisciplinary area, which might in the near future bring considerable clues about Earth's thermal dynamics and formation process.

In this work, we construct an inferential model based on the multigrid priors method to deal, in a generic way, with the geo-neutrino source reconstruction problem. It is an inverse problem; given a region in space V and a finite and small number of measurements of the potential generated on the surface of V by some charge distribution ρ , we try to infer ρ . We present examples of applications and analysis of models in two and three dimensions and we also comment how other a priori information may be included.

Furthermore, we indicate the steps for inferring the best locations for future detectors. The objective is to maximize the amount of information liable to be obtained from experimental measurements. We resort to an entropic method of inference which may be applied right after the results of the multigrid method are obtained.

Geo-neutrinos são anti-neutrinos eletrônicos provindos do decaimento beta de alguns poucos elementos nas cadeias de ^{232}Th e ^{238}U presentes no interior da Terra. Recentes medidas experimentais dessas partículas têm proporcionado grandes expectativas como uma nova maneira de se investigar o interior do planeta diretamente. Trata-se de uma área multidisciplinar nova que poderá no futuro próximo nos trazer grandes esclarecimentos sobre a dinâmica térmica e o processo de formação da Terra.

Neste trabalho, construímos um modelo de inferência baseado no método de multigrid de priors para tratar, de modo genérico, o problema da reconstrução das fontes de geo-neutrinos no interior da Terra. Trata-se de um problema inverso; dada uma região do espaço V e um número finito e pequeno de medidas do potencial gerado na superfície de V por uma distribuição de carga ρ , tentamos inferir ρ . Apresentamos exemplos de aplicações e análises do método em modelos bidimensionais e tridimensionais e também comentamos como outras informações a priori podem ser incluídas.

Além disso, indicamos os passos para se inferir onde detectores futuros devem ser posicionados. O objetivo é maximizar a informação passível de ser obtida das medidas experimentais. Utilizamos um método baseado em inferência entrópica e que pode ser aplicado diretamente depois que os resultados do método de multigrid são obtidos.

Contents

Abstract	3
1 Introduction	7
1.1 Geo-Neutrino Direct and Inverse Problems	8
1.2 Outline	10
2 Preliminary Tools	11
2.1 Probabilities	11
2.1.1 Cox Axioms	11
2.1.2 Bayes in Practice	13
2.2 Entropic Inference	15
2.2.1 Entropy	15
2.2.2 Maximum Entropy	16
2.2.3 Online Updating	16
2.2.4 Experimental Design	17
2.3 General Comments on Inverse Problems	18
3 Tomography Model	21
3.1 Problem Introduction	21
3.2 Multigrid Approach	22
3.2.1 Discretization	23
3.2.2 Estimating Errors	24
3.2.3 Likelihood	25
3.2.4 Priors	26
3.3 Learning Dynamics	27
3.3.1 Parameter Renormalization	29
3.4 Prior Examples	30
3.4.1 Gamma Distributions	30
3.4.2 Reducing Dimensions	31
4 Applications	33
4.1 General Procedure	33
4.2 Toy Model in Two Dimensions	36
4.2.1 Experimental Setup	36
4.2.2 Results and Discussion	37
4.3 Three Dimensional Model	46

4.3.1	Setup	46
4.3.2	Results and Discussion	47
4.4	Geo-Neutrinos?	54
4.4.1	New Detectors	54
5	Conclusions and Future Work	57
A	Geo-Neutrino Oscillation and Spectra	59
A.1	Brief Review of Neutrino Oscillations in Vacuum	59
A.2	Geo-Neutrino Oscillation	62
A.2.1	Interactions with Matter	62
A.2.2	Oscillations in Matter	63
5.2.3	Oscillations inside the Earth	64
5.2.4	Approximations	66
5.3	Spectra	67

1 | Introduction

The idea of using neutrinos as sources of information of Earth's interior was first introduced in the 60's [Eder, 1966, Marx, 1969], only about a decade after the first experimental observations of the particle were made. The subject was then revisited in 1984 [Krauss et al., 1984] as a claim for the development of new detectors that could discriminate sources of electronic anti-neutrinos. However, neutrino physics at that time was not yet developed enough to tackle geo-neutrinos. The solar neutrino flux problem was still considered as open and experimental confirmation of neutrino oscillations and the MSW effect (the change in the effective mass and oscillation parameters due to interactions of neutrinos with matter can lead to oscillation resonance [Wolfenstein, 1978, Mikheyev and Smirnov, 1985]) as a definite solution only came years later with the results of Super-Kamiokande and SNO. Following these developments, the interest in geo-neutrinos grew and their first experimental reports were made in 2005 by the KamLAND Collaboration [Araki et al., 2005]. In 2010 the Italian experiment Borexino also reported strong evidence of observation of neutrinos coming from the Earth's interior [Bellini et al., 2010]. Recently, in 2013, both the KamLAND and Borexino collaborations presented new results following long periods of exposition; Borexino reported the observation of 14.3 ± 4.4 geo-neutrinos over the period of 1353 days [Bellini et al., 2013] and KamLAND reported $116 \pm_{27}^{28}$ events in a total live-time of 2991 days [Gando et al., 2013].

Geo-neutrinos are essentially electronic anti-neutrinos (usually denoted by $\bar{\nu}_e$) originated from the β^- decay process of radioactive elements present inside the Earth. Although many of these different radioactive sources might have been present in early periods of planetary formation, only a small fraction of them are still sufficiently active to contribute to the observed global radioactive power. Those whose half-lives are much smaller than the Earth's age (to date, the oldest minerals analyzed are about $4,404 \pm 0,008$ billion years old and we can use this value as a lower bound on Earth's age [Wilde et al., 2001]) have already decayed into their stable sub-products. On the other hand, elements whose half-lives are much larger than Earth's age might have negligible radioactive powers, depending on their abundance. These are very stringent constraints on the possible sources of geo-neutrinos: we are left only with the decay chains of ^{238}U , ^{235}U and ^{232}Th and also with ^{40}K [Fiorentini et al., 2007]. Another constraint comes from detection limitations. Currently experiments such as Borexino and KamLAND work by observing the inverse beta decay, which has a threshold energy of approximately 1.806 MeV. So geo-neutrinos with energy spectra lying entirely below this limit are invisibles to these detectors. After we carry the process of elimination, only a few nuclei are found to be sources of detectable geo-neutrinos; namely, a few sub-products in the decay chains of ^{238}U and ^{232}Th .

The main geophysical interest on geo-neutrinos comes from the relationship between

the energy they carry away from the interior of the planet and the observed terrestrial heat. The total heat flux from the surface of the Earth is quite well established. Some recent measurements are 47 ± 1 TW [Davies and Davies, 2010] and 46 ± 3 TW [Jaupart et al., 2007]. However, its portion due to radiogenic heat, although believed to play a significant role, is poorly estimated because it relies on the abundance of Earth's radioactive content. Thus, the amount and distribution of these radiogenic heat producing elements inside the mantle are needed for understanding the dynamics of the planet's interior. This includes questions about mantle convection and layer structure, mantle homogeneity, Earth's thermal evolution and the energetics of the core [Fiorentini et al., 2007, Sramek et al., 2013]. Although it is believed that ^{40}K and ^{235}U also play significant roles in heat production, determining only the abundance and distribution of ^{238}U and ^{232}Th by measuring geo-neutrinos may represent a significant step to understand these questions.

The determination of the distribution of ^{238}U and ^{232}Th may also provide some insight on constraining present geochemical models. Although seismic tomography has provided over the last decades valuable information about the density profile of the planet, it is not able to reconstruct the chemical composition of the Earth [Fiorentini et al., 2007]. Also, we are unable to access directly the inner structure of the planet (the deepest hole made so far is about 12 km). So geochemical models often rely on estimates from indirect assumptions, such as chondritic meteorites that resemble the original solar nebula that gave birth to our solar system. There is a whole class of geochemical models, called BSE (Bulk Silicate Earth), which predict slightly different radiogenic heat ratios from different assumptions. A precise measurement of this heat ratio might be able to point out a useful model [Sramek et al., 2013].

Currently only KamLAND and Borexino are able to observe geo-neutrinos due to their low energy threshold. But the increasing interest, especially by geophysicists, is resulting in a number of new experiment proposals [Machado et al., 2012, Autiero et al., 2007, Wurm et al., 2010, Tolich et al., 2007, Barabanov et al., 2009, Learned et al., 2007]. SNO+, which is expected to start taking data soon, will be most likely the third active geo-neutrino detector in the future [Lozza, 2012].

1.1 Geo-Neutrino Direct and Inverse Problems

Our objective throughout this work is to tackle the problem of determining the distribution of geo-neutrino producing elements, such as ^{238}U and ^{232}Th , from measurements obtained on the surface of the Earth. This is the so called inverse problem. But first we will introduce the direct problem of calculating the geo-neutrino number from a source.

Suppose we are given the distribution $\rho(\mathbf{r}, X)$ of some detectable geo-neutrino producing element X inside the Earth and imagine we have M detectors placed on the surface of the globe at points of coordinates $\mathbf{R}_i$, $i = 1, \dots, M$. By measuring the number of events over a period of time, each detector is able to observe $N_{i,X} = N(\mathbf{R}_i, X)$ geo-neutrino candidates coming from $\rho(\mathbf{r}, X)$. Of course $N_{i,X}$ is not a precise quantity; the detectors are subject to many other background effects that mask and mimic the events, including their own efficiency $\epsilon_i(E)$, their volume (or number of protons n_i^p) and their exposure time t_i . The specific type of reaction used by geo-neutrino detectors (normally the inverse beta decay) is also another source of errors. So $N_{i,X}$ is assumed to be always accompanied by

some uncertainty. Now, taking into account the detector's characteristics, we can write $N_{i,X}$ as

$$N(\mathbf{R}_i, X) = n_i^p t_i \int dE \sigma_i(E) \epsilon_i(E) \phi(\mathbf{R}_i, X, E), \quad (1.1)$$

where ϕ is the flux of geo-neutrinos of energy E (sometimes called the differential flux in respect to E and denoted by $d\phi/dE$) and σ_i is the cross section of the reaction used by detector i . ϕ , in turn, is obtained by integrating over the spatial differential flux at the surface of the Earth:

$$\phi(\mathbf{R}_i, X, E) = A_X f_X(E) \int \frac{\rho(\mathbf{r}, X)}{4\pi|\mathbf{R}_i - \mathbf{r}|^2} \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e}(E, |\mathbf{R}_i - \mathbf{r}|) d^3r. \quad (1.2)$$

It represents the number of $\bar{\nu}_e$'s originated from $\rho(\mathbf{r}, X)$ that passes through the surface of the Earth at $\mathbf{R}_i$ with energy E . $f_X(E)$ is the normalized energy spectrum of the produced $\bar{\nu}_e$'s and A_X is the activity of X per unit mass, given by the number of decays per unit time per unit mass of X . The factor $\mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e}$ is the survival probability of the emitted $\bar{\nu}_e$. It represents the probability of an electronic anti-neutrino of energy E to not oscillate into either a tau or muon anti-neutrino during its path from the source to the detector. Its exact expression accounting for matter interactions with the Earth's interior is complicated but can be simplified due to the small and low energy range geo-neutrinos span (~ 2 MeV) and also due to the now known values of the oscillation parameters.

Note that here we are talking about a naturally occurring element X , but what X really represents is an entire decay chain, such as the ^{238}U chain. Once we have the distribution of the head element X , every aspect of the subproducts of its chain can be inferred. Thus, A_X refer specifically to the nuclei X , whereas $f_X(E)$ refers to the energy spectrum of the whole chain, which includes all of its elements.

In order to make (1.1) easier to handle, we substitute the survival probability $\mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e}$ for its average $\langle \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \rangle$ over either the propagation distance $|\mathbf{R}_i - \mathbf{r}|$ or the energy resolution of the detector (see appendix A) and the detector's efficiency ϵ by its average $\langle \epsilon \rangle$ over the small energy interval geo-neutrinos span. Expression (1.1) simplifies to

$$N(\mathbf{R}_i, X) = n_i^p t_i \langle \epsilon_i \rangle \langle \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \rangle \langle \sigma_i \rangle_X A_X \int G(\mathbf{R}_i, \mathbf{r}) \rho(\mathbf{r}, X) d^3r, \quad (1.3)$$

where $G(\mathbf{R}_i, \mathbf{r}) = 1/4\pi|\mathbf{R}_i - \mathbf{r}|^2$ and $\langle \sigma_i \rangle_X$ is the average cross section over the energy spectra [Fiorentini et al., 2007]:

$$\langle \sigma_i \rangle_X = \frac{\int dE \sigma_i(E) f_X(E)}{\int dE f_X(E)}. \quad (1.4)$$

By summing expression (1.3) over the geo-neutrino producing elements X one obtains the total geo-neutrino signal at each detector.

So, given the detectors' characteristics, the quantities related to the decay chain X , our current theory of neutrino oscillations and the distribution $\rho(\mathbf{r}, X)$ of X inside the Earth, we are able to predict the number of geo-neutrinos arriving at each detector up to some degree of experimental uncertainties. But now we turn to the inverse problem, which will be our future subject of investigation: given the M measurements $N(\mathbf{R}_i, X)$

and all the other aforementioned quantities, except for $\rho(\mathbf{r}, X)$, what can we say about $\rho(\mathbf{r}, X)$?

In addition to the currently small number of running geo-neutrino detectors, other experimental difficulties also pose obstacles to the reconstruction of $\rho(\mathbf{r}, {}^{238}\text{U})$ and $\rho(\mathbf{r}, {}^{232}\text{Th})$. At the current experimental level, these are [Fiorentini et al., 2007]:

- Not just only a small number of events can be observed per year but data is usually noisy due to background events;
- Experiments are not able to provide directional information;
- Only about 20% of the events are expected to be due to geo-neutrinos from the ${}^{232}\text{Th}$ chain. Since the total event number is small, it is hard to extract information about the Earth's content of Thorium.

In what follows, we will try to cope with these difficulties in a statistical model of inference.

1.2 Outline

In chapter 2 we show that the most natural way to handle the problem posed at the end of the last section is through the theory of probability. We also mention some useful techniques which we use in our attempt to solve it. In chapter 3 we reintroduce the geo-neutrino inverse problem in a more general way by mentioning only its essential aspects and try to tackle it by developing a set of tomographic techniques based on the tools of chapter 2. Then, in chapter 4, we present some examples of implementations of the techniques of chapter 3 and discuss the results. At the end of chapter 4 we address some comments on applications of the tomography model to geo-neutrinos. Finally, in chapter 5, we make some concluding remarks. In appendix A we review briefly the physics of geo-neutrinos, including the phenomenon of oscillation and the spectra of energy.

2 | Preliminary Tools

This chapter serves only as a summary of the techniques we use in chapter 3 to develop the geo-neutrino inverse problem. More detailed explanations and calculations can be found in [Caticha, 2012]. Other important, but not so modern, references are [Cox, 2001, Jaynes, 2003].

2.1 Probabilities

The primary mathematical tool for dealing with problems of incomplete information is the theory of probabilities. But is it the only one? If not, are there better tools? And more importantly, to start off, how do we know the validity of the first statement? All these questions can be answered once we decide what we mean by “inference”.

Our objective, unlike deductive logic, is to be able to say something about an assertion “A” in the view of a statement “B”, even when “A” does not follow, in the logical sense, from “B”. To be direct to the point, what we want is to address the plausibility (the degree of trueness, or degree of belief) of “A” in the view of “B”. We denote this by the symbol $(A|B)$.

We start by a set of general principles that mathematically translates our desires into manipulation rules of the beliefs $(A|B)$. Their role is to establish a mathematical framework and to rule out of the theory undesirable situations, such as giving different degrees of belief, in the absence of any other information, to different assertions we know to be true, giving different degrees of belief to two different representations of the same assertion and so on. In any case, it is still possible to include the aforementioned cases by removing some of the axioms. However, the emerging theory will not be useful scientifically as it will allow manifestly inconsistent situations. The main result we obtain, after analyzing some particular cases, is that our theory of inference must be identical to the standard theory of probability.

2.1.1 Cox Axioms

The first axiom deals with our desire to quantify the degrees of belief and to be able to order them. That is, we want our theory to include transitivity: if $(A|B) > (A|C)$ and $(A|C) > (A|D)$, then $(A|B) > (A|D)$.

I. The degree of belief $(A|B)$ is represented by a real number.

Secondly, we require consistence when calculating the degree of belief of two different representations of the same assertion. For instance, we require associativity in both sum

and product of assertions under the operation of $(\cdot|\cdot)$, so that $(A(BC)|D) = ((AB)C|D)$ and $((A+B)+C|D) = (A+(B+C)|D)$.

- II. If the degree of belief in “A” can be computed in two different ways, the results must be the same.

Now we look at specific cases of consistence. First we impose the desire that all truths and untruths should be equally plausible.

- III. For every “A”, $(A|A) = v_v$ and $(\bar{A}|A) = v_f$,

where both constants are, a priori, unknown and not necessarily equal to 0 and 1. Secondly, we impose that the plausibility of composite assertions, such as “AB” and “A+B”, must be decomposable into simpler ones, involving just “A” or “B”.

- IV. The degrees of beliefs $(AB|C)$ and $(A+B|C)$ must be related to some subset of the separate degrees of beliefs $\{(A|C), (B|C), (A|BC), (B|AC)\}$ by functions f and g , respectively.

Note that if “A” and “B” are mutually exclusive events in the view of “C”, $(A+B|C)$ is function of $(A|C)$ and $(B|C)$ only, so f must have at most two arguments. So now if “A”, “B” and “C” are mutually exclusive in the view of “D”, together with (II) and (IV), group associativity follows: $((A+B)+C|D) = (A+(B+C)|D) \Rightarrow f(f((A|D), (B|D)), (C|D)) = f((A|D), f((B|D), (C|D)))$. That is, $f(f(x, y), z) = f(x, f(y, z))$.

Manipulation Rules

We now regraduate f according to $f(x, y) = \phi^{-1}(\phi(x) + \phi(y))$. Note that if ϕ is crescent the order of the numbers x do not change. For mutually exclusive events “A” and “B”, the direct consequence is that $\phi(A+B|D) = \phi(A|D) + \phi(B|D)$. In the general case, we obtain

$$\phi(A+B|C) = \phi(A|C) + \phi(B|C) - \phi(AB|C) . \quad (2.1)$$

A more complicated argument, but which follows the same ideas of constraining the possible outcomes from particular situations and then regraduating the obtained result gives

$$P(AB|C) = P(A|BC)P(B|C) , \quad (2.2)$$

and the sum rule becomes simply

$$P(A+B|C) = P(A|C) + P(B|C) - P(AB|C) . \quad (2.3)$$

This particular choice of scale also gives $P(A|\bar{A}) = 0$ and $P(A|A) = 1$. With (2.3), we also conclude that $P(A|D) + P(\bar{A}|D) = P(A+\bar{A}|D) = 1$

Equations (2.3) and (2.2) are readily recognized as the usual sum and product rules of probability theory. Together with the other mentioned results that follow directly from the axioms after proper regraduation, they allow us to completely identify what we called plausibilities to the usual probabilities. We thus no longer make reference to “plausibility” or “degrees of belief” in the context of inference, but to probabilities.

Bayes' Theorem

Given two assertions “A” and “B” on the view of some hypothesis “C”, the product rule (2.2) states that their joint probability can be written as

$$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C) . \quad (2.4)$$

Thus, we find that the conditional probability of “A” on “B” is proportional to the conditional probability of “B” on “A”, a result known as Bayes' Theorem.

2.1.2 Bayes in Practice

Suppose now that Y represents a set of data collected by an observation and X some quantity that we happen to know the probability if I is given. I represents all the relevant background information to the process of inference about X ; that is, all our available information concerning X , prior to Y , is included in I . The problem we face now is to determine how the probability of X should change once we have Y available. More specifically, we want to update from the *prior* distribution $P(X|I)$ to some *posterior* distribution $Q(X|I)$ that incorporates the new data.

However, in order to proceed, we must know how the quantities X and Y relate to each other. We will at this stage assume that the *likelihood* distribution $P(Y|XI)$ is known and we will postpone its discussion. Now, (2.4) states that if we want to be consistent with our theory of inference, we must take

$$P(X|YI) = P(X|I) \frac{P(Y|XI)}{P(Y|I)} . \quad (2.5)$$

The denominator, called *evidence*, is obtained from the normalization of $P(X|YI)$,

$$P(Y|I) = \int dX P(X|I) P(Y|XI) , \quad (2.6)$$

and is usually one of the main sources of difficulties in practical applications (normally due to the high dimensionality of X). To connect this result with the updated distribution $Q(X|I)$, we simply take $Q(X|I) = P(X|YI)$, agreeing that we should only update the probabilities of X if some new acquired data Y says we have to (we do not use information we do not have). For now on we will drop the Q notation and denote the posterior also by P . We will distinguish each factor by its arguments, just as in (2.5).

Prior

The prior distribution $P(X|I)$ is constructed based on everything we happen to know about X before we look at Y . The specific family of distributions used to describe $P(X|I)$ depends on the nature of the problem and is usually specified by constraints (such as positiveness of X , a known value for $\langle X \rangle$ and so on). However, some general remarks can be made concerning its relationship to the likelihood in (2.5).

First, suppose $P(X|I)$ is a very narrow distribution in comparison to $P(Y|XI)$ around some value $\langle X \rangle$. This represents a situation where we have a high expectation that X only takes values near $\langle X \rangle$ and the new observed data brings us very little information

about X . The conclusion is that the posterior distribution will be very much similar to $P(Y|XI)$.

Now, suppose the opposite: $P(X|I)$ is very fat-tailed in comparison to $P(Y|XI)$, a situation that represents our ignorance about X in the absence of data or in the presence of noisy data. In this case, almost every information contained in $P(Y|XI)$ will be incorporated into the posterior, which will now behave similarly to the likelihood.

Both cases can be pitfalls when assigning priors. In the next chapter, we will investigate these situations and see how prejudice and complete ignorance may lead to undesirable effects in the process of inference.

Likelihood

The likelihood distribution $P(Y|XI)$ informs us how probable it is to obtain a measurement Y assuming that the given value of X and all the background information are correct. Because any type of measurement is subject to experimental errors, the likelihood is usually distributed around some region and not simply a Dirac delta of some particular value of Y . However, the more accurate the collected data Y is, the more narrow the likelihood will be, and the opposite is also true. This is represented by the experimental error ξ , which in the absence of any other uncertainty from I serves as a measure of the variance of $P(Y|XI)$.

As an example which we will reproduce later, suppose we are given a set of independent measurements $\{(\eta_i, t_i)\}$ of some physical quantity η parametrized by t . Imagine that we also have as background information a model $M : \eta_i = \alpha t_i + \xi_i$, linking both quantities by a constant α with the addition of some noise ξ_i . For instance, η could represent the velocity of an e charged particle of mass m , placed at rest in a region of constant electric field E , so that the theory would predict α to be Ee/m . Assuming ξ is normally distributed around zero with variance σ ,

$$P(\{\eta_i, t_i\}|\alpha I) = \int d\xi P(\xi|\{\eta_i, t_i\}I) P(\alpha|\xi\{\eta_i, t_i\}I) \propto \prod_i \exp -\frac{1}{2} \left(\frac{\eta_i - \alpha t_i}{\sigma_i} \right)^2, \quad (2.7)$$

which follows directly from the fact that $P(\alpha|\xi\{\eta_i, t_i\}I) = \prod_i \delta\left(\frac{\eta_i - \xi_i}{t_i} - \alpha\right)$. Were the error distributed by another family of distributions, the functional form of the likelihood would change. We will justify the choice of the normal distribution later in the next sections.

Assuming a constant prior, the posterior is given by (2.7) and one possible experimental estimative for α is the value that maximizes the likelihood:

$$\alpha_{\text{ML}} = \arg \max_{\alpha} P(\{\eta_i, t_i\}|\alpha I) = \arg \max_{\alpha} \sum_i \left(\frac{\eta_i - \alpha t_i}{\sigma_i} \right)^2, \quad (2.8)$$

a method known as least-squares fitting. If, on the other hand, the prior is not constant, but is distributed around the likelihood, we have to take it in account if we want to make the same estimative for α :

$$\alpha_{\text{MAP}} = \arg \max_{\alpha} P(\{\eta_i, t_i\}|\alpha I) P(\alpha|I), \quad (2.9)$$

which is known as the maximum a posteriori estimative. However, both estimates of α are pointwise and hide any uncertainty that we might have in its true value, as expressed in the likelihood. To be more consistent with our lack of certainty, we may take as an estimative of α its mean value over the posterior distribution

$$\langle \alpha \rangle = \int d\alpha P(\alpha | \{\eta_i, t_i\} I) \alpha, \quad (2.10)$$

along with its uncertainty $\sigma^2 = \langle \alpha \rangle^2 - \langle \alpha^2 \rangle$. This way, we avoid at some extent the arbitrariness of choosing a representative for α in the lack of information or data.

2.2 Entropic Inference

Bayes' Theorem is a tool for updating probabilities in the view of new data but is unable to handle other types of available information. However, as we will see, it can be deduced as a special case of a more general method of updating probabilities, the Maximum Entropy. We devise a functional S that, when maximized, chooses the best probability density $P(x)$ in respect to a given prior $Q(x)$ and a set of constrains, according to some previously chosen design criteria. Thus, from the point of view that information is anything which forces an update in one's probabilities, the framework that we will describe next is the most general way we have to handle inference.

2.2.1 Entropy

The construction of the entropy functional S for continuous random variables follows a similar path to the one we made when devising the theory of probability as the correct way to handle degrees of beliefs. We establish some general criteria which we want our theory to reflect, translate these desires mathematically in the form of axioms and then we develop, with the assistance of some particular cases, their consequences. Different design criteria will lead to different functionals S . Here we are interested in a particular one that takes two distributions, a prior Q and a posterior P , and rank them according to the following criteria:

- I. Locality: In absence of new information about some $\mathcal{D} \subseteq \mathcal{X}$, $P(X|\mathcal{D}) = Q(X|\mathcal{D})$, for $X \in \mathcal{X}$.

Which simply states that if we have no new information of a subsystem, then we simply maintain the prior information as the posterior.

- II. Invariance under coordinate change: The transformed probability densities $P'(Y)$ and $Q'(Y)$ induced by a coordinate transformation $X = \Gamma(Y)$ will generate the same ranking as $P(X)$ and $Q(X)$.

As it is known, change of variables under the integral sign transforms probability densities $P(X)$ and $Q(X)$ into $P'(Y) = P(X)J(Y)$ and $Q'(Y) = Q(X)J(Y)$, where J is the Jacobian. We impose that the ranking of distributions $Q'(Y)$ in relation to $P'(Y)$, given by $S[P', Q']$, must be the same as the one given by $S[P, Q]$. Their numerical values do not necessarily need to coincide.

- III. Independence: If two systems are believed to be independent, so that the joint prior is described by $Q_1(X_1)Q_2(X_2)$, and new information induces $Q_1(X_1) \rightarrow P_1(X_1)$ and $Q_2(X_2) \rightarrow P_2(X_2)$, then $S[P, Q]$ must update $Q_1(X_1)Q_2(X_2)$ to $P_1(X_1)P_2(X_2)$.

The consequence of carrying the analysis of these criteria (see [Caticha, 2012]) is that distributions P must be ranked in relation to the prior Q according to

$$S[P|Q] = - \int dX P(X) \log \frac{P(X)}{Q(X)}, \quad (2.11)$$

and subject to any constraints we have to impose. As a (not so direct) consequence, Bayes' rule of updating priors (2.5) can be deduced as a particular case of maximization of (2.11).

2.2.2 Maximum Entropy

Suppose we want to choose from a subset of all the existing probability density functions one that best recognizes some previously specified prior distribution and, simultaneously, respects the basic criteria imposed in section 2.2.1. That is, given the functional S of equation (2.11) and a set of constraints $\{C\}$, we want to find the distribution P that best represents Q . We simply look at the ranking of distributions created by S and choose the one P sitting on the top of it by maximizing S subject to $\{C\}$.

Entropy maximization in this point of view is simply a tool for updating prior distributions when information is available. It handles not only constraints, but also information in the form of data by having Bayes' theorem included.

As an example, suppose we have no prior knowledge about how some variable ξ should be distributed, so that its prior is inferred to be a constant P_0 . But imagine now that we receive the information that actually ξ has some expected value μ and variance σ :

$$\int d\xi P(\xi) = 1, \quad \int d\xi P(\xi) \xi = \mu, \quad \int d\xi P(\xi) \xi^2 = \mu^2 + \sigma^2. \quad (2.12)$$

Entropy maximization subject to these constraints reads

$$\delta \left\{ - \int d\xi P(\xi) \log \frac{P(\xi)}{P_0} + \lambda_1 \left[\int d\xi P(\xi) - 1 \right] + \lambda_2 \left[\int d\xi P(\xi) \xi - \mu \right] + \lambda_3 \left[\int d\xi P(\xi) \xi^2 - \mu^2 - \sigma^2 \right] \right\} = 0, \quad (2.13)$$

and we conclude that $P = \exp(1 - \lambda_1 - \lambda_2 x - \lambda_3 x^2)$, or, by choosing the Lagrange multipliers λ_i to satisfy the constraints (2.12), $p = \exp -\frac{1}{2\sigma^2}(\xi - \mu)^2 / \sqrt{2\pi\sigma^2}$, which is the normal distribution we assumed in (2.7).

2.2.3 Online Updating

Sometimes the posterior distribution obtained from a Bayes step has some complicated form. Specially when the prior and the likelihood belong to distinct families of distributions, the posterior obtained can have some undesirable functional form. One way to

simplify the model's implementation is to approximate the posterior $P(X|Y)$ by some simpler distribution $Q(X)$, which is usually taken to belong to the family of distributions of the prior $P(X)$. We first decide the exact functional form of $Q(X)$ we want to approximate $P(X|Y)$ into and then we project $P(X|Y)$ onto the space of $Q(X)$'s by maximizing their relative entropy. We therefore seek a $Q(X)$ that best approximates $P(X|Y)$ in the entropic sense.

The idea of first obtaining a posterior distribution from Bayes' Theorem and then projecting it onto the space of distributions where the prior lives is usually referred as online learning or online updating.

Note that in practice we use exchanged roles for $P(X|Y)$ and $Q(X)$ in contrast to the idea of projecting a prior into a posterior. $P(X|Y)$ is viewed as the posterior distribution which we want to reproduce, although it is known. $Q(X)$ is viewed as the prior:

$$S[P|Q] = - \int dX P(X|Y) \log \frac{P(X|Y)}{Q(X)} . \quad (2.14)$$

Next we maximize (2.14) in the parameters θ of $Q(X)$ to obtain explicitly the projected distribution.

This is very useful to construct mean-field approximations of complicated posteriors. For instance, suppose Bayes' Theorem gives us $P(X|Y) = P(X)P(Y|X)/P(Y)$. If $Q(X) = \prod_i Q(x_i|\theta_i)$ is our intended mean-field approximation to $P(X|Y)$, and $P(X) = \prod_i P(x_i|\eta_i)$, then

$$S[P|Q] = - \int dX P(X|Y) \left(\log P(Y|X) - \sum_i \log \frac{Q(x_i|\theta_i)}{P(x_i|\eta_i)} \right) . \quad (2.15)$$

Maximization in relation to θ_j yields

$$\frac{\partial S}{\partial \theta_j} = - \int dX P(X|Y) \frac{1}{Q(x_j|\theta_j)} \frac{\partial Q(x_j|\theta_j)}{\partial \theta_j} = 0 . \quad (2.16)$$

That is,

$$\left\langle \frac{\partial Q(x_j|\theta_j)/\partial \theta_j}{Q(x_j|\theta_j)} \right\rangle_{P(X|Y)} = 0 . \quad (2.17)$$

By solving (2.17) we obtain the parameters θ_i and, consequently, $Q(X)$.

2.2.4 Experimental Design

Suppose we have some prior knowledge Y_0 , in the form of data, from which we, somehow, manage to infer that $Q(X|Y_0)$ is the probability density which best describes X at the level of Y_0 . Now, suppose we want to go further and obtain a new set of data Y to improve our description of X . How should it be gathered?

This question can be stated in a more practical way: from a list of possible distributions $P(X|Y_0, Y, I_\epsilon)$, ϵ labeling the experimental setup, which one will be the most informative, given $Q(X|Y_0)$? This is a decision problem. In order to investigate it, we must first decide what "informative" means.

We will resort to the interpretation of $-S[P|Q]$ as a measure of the information gained by going from a prior Q to a posterior P . Therefore, “the most informative” here refers to the one which maximizes $-S[P|Q]$. This quantity is often called the mutual information between P and Q . However, we do not know Y because it is not available to us before it is gathered, so the quantity which we must maximize is rather the expected value of $-S[P|Q]$ over Y [Loredo, 2003, Caticha, 2012]:

$$-\langle S[P|Q] \rangle_\epsilon = \int dY P(Y|Y_0, I_\epsilon) \int dX P(X|Y_0, Y, I_\epsilon) \log \frac{P(X|Y_0, Y, I_\epsilon)}{Q(X|Y_0)} . \quad (2.18)$$

Note that the maximization of (2.18) is carried in the parameter ϵ , which represents some experimental setup variable which we are allowed to vary. The quantity $P(Y|Y_0, I_\epsilon)$ can be obtained by marginalizing over X :

$$P(Y|Y_0, I_\epsilon) = \int dX' P(X'|Y_0, I_\epsilon) P(Y|X', Y_0, I_\epsilon) . \quad (2.19)$$

2.3 General Comments on Inverse Problems

Generally speaking, inverse problems are concerned with the determination of the causes X of an observed effect Y . The direct problem, on the other hand, is concerned with calculating the effects Y of a given cause X . They are related in the sense that the quantities Y and X are linked, in the direct problem, by some function $\mathbf{G} : X \mapsto Y$ which represents a theory; however, the inverse problem is usually much more complicated as it involves the inversion of $\mathbf{G}$ in some way. To make matters worse, in practical applications, we usually have to deal not only with a finite number of measurements Y but also with measurement noises ξ :

$$Y = \mathbf{G}(X) + \xi . \quad (2.20)$$

Note that not only this is a problem of inference, but Bayes’ Theorem gives right away the inversion of (2.20):

$$P(X|Y) \propto P(Y|X) . \quad (2.21)$$

Given a set (Y, ξ) , we may have to cope with three distinct types of difficulties: first if a solution X exists, second if X is unique and third if X depends continuously on Y . The two first conditions refers to the existence of an inverse of $\mathbf{G}$. The third one is concerned with the stability of the solution X (if small changes in Y arbitrarily modifies X). If any of these is not met, the inverse problem is said to be ill-posed.

In the linear case of finite dimensions, which will be our main concern here, $\mathbf{G}$ becomes a finite dimensional linear transformation $N \times M$. In this framework, (2.20) may have two distinct behaviors: first when $N \geq M$ and second when $N < M$.

In order to understand the general behavior of the inversion of (2.20) in each of these cases, we must analyse the properties of the inverse of $\mathbf{G}$. To do that, we decompose $\mathbf{G}$ in terms of the eigenvectors of $\mathbf{G}^T \mathbf{G}$ and $\mathbf{G} \mathbf{G}^T$, $\mathbf{v}_k$ and $\mathbf{u}_k$ respectively,

$$\mathbf{G} = \sum_{i=1}^r s_i \mathbf{u}_i \mathbf{v}_i^T , \quad (2.22)$$

where each of these eigenvectors is associated with one of the $r \leq \min(N, M)$ positive eigenvalues s_i^2 . This is called the singular value decomposition (SVD) of $\mathbf{G}$ and the quantities s_i are called singular values. If we arrange every $\mathbf{u}_k$ and $\mathbf{v}_k$ as the columns of matrices $\mathbf{U}$ and $\mathbf{V}$, respectively, we may write (2.22) as $\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{S}$ can assume the forms:

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{0} \end{pmatrix}, \quad (\mathbf{D} \quad \mathbf{0}), \quad (\mathbf{D}), \quad (2.23)$$

with $\mathbf{D} = \text{diag}(s_1, \dots, s_r)$, depending on N and M . Regarding the first two conditions of ill-posedness, if $N \geq M$ (so that $r \leq M$) or if $N < M$ (so that $r \leq N$), we see from (2.23) two types of behaviors. Respectively, the first one is that a solution X may not exist if there are errors and we do not recognize their existence. The second one, is that, if the system is undetermined, the solution X is not unique.

The SVD generalizes the notion of inverse of a matrix (also known as the pseudo-inverse). We can attempt to solve the inverse system by writing

$$X = \mathbf{G}^{-1}Y = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T Y = \sum_{i=1}^r \frac{Y_{i,u}}{s_i} \mathbf{v}_i, \quad (2.24)$$

where $Y_{i,u}$ is the projections of Y along the vectors u_i . Small singular values s_i are associated with the process of amplification of noise. For instance, suppose we perturbate Y with an error $\delta Y = \delta y \mathbf{u}_k$ along the k eigenvector $\mathbf{u}_k$. It produces a perturbation $\delta x = \delta y / s_k$ along $\mathbf{v}_k$. We see that if s_k is small, the resulting perturbation in X will be large.

When analysing the constructed likelihood distribution in our applications, we will often find very small singular values. In the normal distribution case, the effect of the noise amplification is to induce large variances along some directions.

3 | Tomography Model

Our objective in this chapter is to reformulate the inverse problem of geo-neutrinos of section 1.1 in a general way and to propose a Bayesian tomography model for it. In terms of the geo-neutrino problem, the method we develop has the purpose of reconstructing the bulk distribution of radioactive elements only from the non-directional measurements of the emitted particles. We will start with a generalization of equation (1.3), try to develop a multigrid model to reconstruct ρ and later we will discuss prior information.

3.1 Problem Introduction

Suppose we have a set of measurements $\mathbf{N} = \{N_i\}_{i=0,\dots,M}$ over a physical system and that some theory establishes a linear relationship between $\mathbf{N}$ and some quantity f , function of variables χ . We may generally describe $\mathbf{N}$ in the form

$$\mathbf{N} = \int_V \mathbf{G}(\chi) f(\chi) d\chi, \quad (3.1)$$

where $\mathbf{G}$ is a list of Green functions which depends on the variables χ defined over a region V . The vector notation of (3.1) makes it explicit that each component $\mathbf{G}$ depends on the corresponding N_i . For instance, χ may represent spacial coordinates $\mathbf{r}$, V some region in space where f is defined at and each N_i be obtained at a different point $\mathbf{r}_i$ of space. In that case, $\mathbf{G}$ may be some function that depends on the distance $|\mathbf{r} - \mathbf{r}_i|$, so that the dependence on the index i is spatial. In chapter 4, our examples of applications will follow that line.

The choice of a specific $\mathbf{G}$ is part of the information given by the theory behind our subject of investigation. If we are dealing with charge distributions and measuring the electric potential, the correct choice for $\mathbf{G}$ would be the electromagnetism's Green function. On the other hand, if $\mathbf{N}$ represents the number of particles flowing through some surface just like (1.3) ¹, $\mathbf{G}$ should be inversely proportional to the square of the distance between the points of production and detection.

Given $\mathbf{N}$ (and its corresponding experimental errors), $\mathbf{G}$ and any other information $\mathbf{I}$ we might have access to, but which is not incorporated by $\mathbf{G}$, our aim is to infer whatever information is available regarding $f(\chi)$. That is, we want to reconstruct the original $f(\chi)$ which generated $\mathbf{N}$. Our attempt to solve the problem is divided into three central sub-problems.

¹For the details concerning the geo-neutrino case, see appendix A

The first sub-problem is to decide how to represent the region V and the variables χ in a discretized framework. It does not arise merely because we are interested in practical applications and so we must be able to represent V computationally, but because we have a finite (and very likely small) number of measurements M . As explained in section 2.3, if we attempt to invert (3.1) with M much smaller than the number of variables χ_i that we chose to represent V , then we will not be able to say much about each $f(\chi_i)$. On the other hand, if the number of points χ_i is much smaller than M , we might be wasting information we already have in hand about $f(\chi)$.

Which specific type of discretization to use is also a non-trivial issue. By choosing a set of points χ_i over another to represent V , we might be introducing some bias into the inference process. To be more clear of what we are talking about, suppose we are trying to reconstruct some distribution in V by measuring some potential on its surface. What will happen if we represent some sub-region U of V in a coarser way for no good reason? Will it generate a more accurate reconstruction of $f(\chi)$ than if we had represented U in a more refined manner? The resulting process of inference of $f(\chi)$ in these two examples will be different, and that difference is what we mean by “bias”.

Once we have decided how we are going to treat the discretization of V , we can construct the likelihood distribution in the same way we have done in the example of section 2.1.2. This step leads us to our second sub-problem: how should we incorporate the available information into our model? More specifically, we must choose from the bookshelf of Bayesian learning algorithms one that fits our necessities and relates well with our discretization method.

But how do we handle prior information? It can enter the model in many ways. For instance, we can choose some biased discretization of V , as we have said earlier, representing some type of prior information. But we are also interested in choosing a specific family of distributions to represent our initial knowledge of each $f(\chi_i)$. This is the third sub-problem we are faced with.

3.2 Multigrid Approach

Our approach to the first problem is based on the multigrid priors method developed in [da Rocha Amaral et al., 2004, da Rocha Amaral et al., 2007, da Silva Barbosa, 2011]. The idea is to represent V by some coarse initial grid and to iteratively refine it until we reach some limit of resolution posed by the experimental data. At each iteration, we first extract from the data all the information it contains about the $f(\chi_i)$ ’s. Then we propagate the obtained information to the next scale, which will be a finer approximation to V . Therefore, the grid representation at some scale d must be constructed so to represent a slicing of V at $d - 1$.

More specifically, we construct a set of lattices $\{\Lambda_d\}_{d=0,\dots,D}$ to represent V at degrees of resolution indexed by d , such that a region volume Δv_k^d is associated to every cell $k \in \{\Lambda_d\}$. We assume that for a fixed d these regions are non-overlapping, so that the sum of their volumes equals the original volume of V . As a consequence of this particular type of grid construction, any lattice region $k \in \Lambda_{d-1}$ can be obtained by blocking together a set of cells $j(k) \in \Lambda_d$. Figure 3.1 illustrates the process for a particular square lattice.

Proceeding, we now have to decide how to discretize the integrand of (3.1) and see

Figure 3.1: Grid refinement process for a particular square region.

how well the resulting sum approximates the integral of (3.1).

3.2.1 Discretization

Let the integrand of (3.1) be $\mathbf{H}(\chi) = \mathbf{G}(\chi)f(\chi)$ and suppose for now that we are trying to solve the forward problem, where $\mathbf{H}(\chi)$ is given. At each scale of resolution d we discretize $\mathbf{H}(\chi)$ by approximating it with a set of step functions over the grid cells. That is, we assign constant quantities $\mathbf{H}_i^d$ to every grid cell i at each scale d and let the process of refinement of the grid dictate the quality of the approximation. The first approximations may be quite rude, as seen in figure 3.1, but, as the initial grid becomes more and more subdivided, the more accurate the approximation becomes.

The first step is to write $\mathbf{H}$ in terms of characteristic (or indicator) functions $\mathbf{1}_k^d(\chi)$ over the grid cells:

$$\mathbf{H}(\chi) = \sum_{k \in \Lambda_d} \mathbf{1}_k^d(\chi) \mathbf{H}(\chi) = \sum_{k \in \Lambda_d} \mathbf{H}_k^d(\chi). \quad (3.2)$$

Next, we expand each $\mathbf{H}_k^d$ in series around some fixed point $\bar{\chi}_k$, so that (3.1) reads, at the first order,

$$\mathbf{N} = \sum_{k \in \Lambda_d} \int_{\Delta v_k^d} \mathbf{H}_k^d(\chi) d\chi \simeq \sum_{k \in \Lambda_d} \int_{\Delta v_k^d} \left(\mathbf{H}_k^d(\bar{\chi}_k) + \sum_j \frac{\partial \mathbf{H}_k^d}{\partial \chi_j} \bigg|_{\chi=\bar{\chi}_k} (\chi'_j - (\bar{\chi}_k)_j) \right) d\chi', \quad (3.3)$$

where the index j runs throughout the components of χ . $\bar{\chi}_k$ is the representative point of cell k . Note, however, that the only constraint imposed on each $\bar{\chi}_k$ is that it must lie inside the cell of volume Δv_k^d and represented by the index k . Apart from that restriction, we are free to choose any point inside Δv_k^d . This arbitrariness rises the question: which point should we choose? In the forward model, since we know $\mathbf{H}(\chi)$, the best choice is clearly the one which minimizes the numerical error of approximation (3.3), something similar to the midpoint rule of the rectangle method. But now, back to the inverse problem, how do we decide which point to choose if we do not know $\mathbf{H}(\chi)$?

Different choices of $\bar{\chi}^d$, the list of all $\bar{\chi}_k$ at scale d , may result in very different inferred values for $f(\bar{\chi}_k)$ if d is small. We handle this arbitrariness by including a discretization error $\boldsymbol{\eta}$ into (3.3). If we further approximate each $\mathbf{H}_k^d$ to the zeroth order, the first term of (3.3) gives us the variables of the linear system and the second term an estimative of

the discretization error:

$$\mathbf{N} \simeq \sum_{k \in \Lambda_d} \mathbf{H}_k^d(\bar{\chi}_k) \Delta v_k^d + \boldsymbol{\eta}. \quad (3.4)$$

Note that the components of $\boldsymbol{\eta}$ in (3.3) depend on the measurement index i and on the set of points $\bar{\chi}^d$ at which we expanded $\mathbf{H}_k^d$. Therefore, at each degree of resolution and for each detector we may have different discretization errors η . However, $\boldsymbol{\eta}$ is rather incalculable in practice. Again, that is because we do not know $f(\chi)$ prior to the process of incorporating the information from the data into our model. So how do we estimate it in practice? We will go back to this issue later and just assume that for now it is given.

3.2.2 Estimating Errors

Now, each component of $\mathbf{N}$ is a number obtained by some measurement process over our system of interest and involves experimental uncertainties. In order to obtain a statistical model linking (3.4) consistently to the real data $\mathbf{N}$, we must also take into account measurement noises $\boldsymbol{\epsilon}$ and rewrite (3.4) as

$$\mathbf{N} \simeq \sum_{k \in \Lambda_d} \mathbf{H}_k^d(\bar{\chi}_k) \Delta v_k^d + \boldsymbol{\eta} + \boldsymbol{\epsilon} = \sum_{k \in \Lambda_d} \mathbf{G}_k^d(\bar{\chi}_k) f_k^d(\bar{\chi}_k) \Delta v_k^d + \boldsymbol{\xi}, \quad (3.5)$$

where $\boldsymbol{\xi} = \boldsymbol{\epsilon} + \boldsymbol{\eta}$. As a matter of simplification, we will omit the dependence of $\mathbf{G}_k^d(\bar{\chi}_k)$ and $f_k^d(\bar{\chi}_k)$ on $\bar{\chi}_k$ and we will also absorb the volumes Δv_k^d either into $\mathbf{G}_k^d$ or into f_k^d , depending on the specific problem posed. Taking these notations into account and rewriting (3.5) in matrix form, we have

$$\mathbf{N} \simeq \mathbf{G}^d \mathbf{f}^d + \boldsymbol{\xi}. \quad (3.6)$$

The error $\boldsymbol{\epsilon}$ is characteristic of the detectors (or the processes of measurement) used to gather $\mathbf{N}$. Thus, assuming the detectors are independent, each component of $\boldsymbol{\epsilon}$ should be expected to have some associated probability distribution $P(\epsilon_i | I_i)$, where I_i represents all the known information about detector i . We may assume, for instance, that all we know about the noise ϵ_i is that it is distributed around the origin and has a known variance σ_{ϵ_i} . In that case, as we have shown in section 2.2.2, maximum entropy states that ξ_i should be normally distributed around zero with variance σ_{ξ_i} :

$$P(\epsilon_i | I_i) = \mathcal{N}(0, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_{\epsilon_i}^2}} \exp -\frac{1}{2\sigma_{\epsilon_i}^2} \epsilon_i^2. \quad (3.7)$$

The joint distribution $P(\boldsymbol{\epsilon} | \mathbf{I})$ of all data gathered is obtained by taking the product of each $P(\epsilon_i | I_i)$ and can be written in matrix notation as

$$P(\boldsymbol{\epsilon} | \mathbf{I}) = \frac{1}{\sqrt{2\pi |\boldsymbol{\Sigma}_{\epsilon}^{-1}|^2}} \exp -\frac{1}{2} \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \boldsymbol{\epsilon}. \quad (3.8)$$

However, what we really need to estimate is the value of $\boldsymbol{\xi}$, which involves $\boldsymbol{\eta}$. As it was said earlier, the unknown character of $\boldsymbol{\eta}$ lies on our lack of knowledge of $f(\chi)$, so we cannot just calculate the error in (3.3) and use its exact value as a measure of the error. But if we could assume that $\boldsymbol{\eta}$ is also normally distributed around zero with some variance

$\Sigma_\eta^{-1},$

$$P(\boldsymbol{\eta}) \propto \exp -\frac{1}{2} \boldsymbol{\eta}^T \Sigma_\eta^{-1} \boldsymbol{\eta}, \quad (3.9)$$

then we would have our problems translated into Σ_η^{-1} and so $\boldsymbol{\xi} = \boldsymbol{\epsilon} + \boldsymbol{\eta}$ would also be normally distributed:

$$P(\boldsymbol{\xi}) = \int P(\boldsymbol{\epsilon}, \boldsymbol{\xi}) d\boldsymbol{\epsilon} = \int P(\boldsymbol{\xi} | \boldsymbol{\epsilon}) P(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \propto \exp -\frac{1}{2} \boldsymbol{\xi}^T \Sigma_\xi^{-1} \boldsymbol{\xi}, \quad (3.10)$$

where $\Sigma_\xi^{-1} = \Sigma_\epsilon^{-1} \Sigma_\eta^{-1} / (\Sigma_\epsilon^{-1} + \Sigma_\eta^{-1})$. The value of Σ_η^{-1} is still unknown to us, though, and requires some type of estimation. What we have gained with this assumption is that now we have a simple way to express the distribution of the joint error $\boldsymbol{\xi}$, which will allow us to construct the likelihood distribution in the same way we did in section 2.1.2.

One possible way to proceed is to assume that Σ_η^{-1} decreases with some power of the grid size, just as in the method of integration by rectangles. In this case, from (3.3), we can estimate $\boldsymbol{\eta}$ being proportional to $n(d)^{-2/s}$, where $n(d)$ is grid d's size and s is the dimension of V . We then may take Σ_η also to be $\gamma n(d)^{-2/s}$, γ being the proportionality factor which still remains free.

3.2.3 Likelihood

Let now all the information given by the set of matrices $\{\mathbf{G}^0, \dots, \mathbf{G}^D\}$ and $\mathbf{I}$, the assumed background information of section 3.1, be $\mathbf{I}^D$. D here denotes the finer scale we plan to achieve in the multigrid method. As we have said earlier, our final aim is to obtain from some Bayesian learning algorithm, which involves Bayes' Theorem, the posterior distribution $P(\mathbf{f}^D | \mathbf{N}, \mathbf{I}^D)$. However, the multigrid method is iterative and it requires us to obtain the posteriors $P(\mathbf{f}^d | \mathbf{N}, \mathbf{I}^D)$ at every scale d in order to obtain $P(\mathbf{f}^D | \mathbf{N}, \mathbf{I}^D)$. For that purpose, we must construct likelihood distributions $P(\mathbf{N} | \mathbf{f}^d, \mathbf{I}^D)$ for each $d \leq D$. This is done exactly as in section 2.1.2, using equations (3.6) and (3.10). First we marginalize $P(\mathbf{N} | \mathbf{f}^d, \mathbf{I}^D)$ over $\boldsymbol{\xi}$ and use the product rule to write the integrand as $P(\mathbf{N} | \boldsymbol{\xi}, \mathbf{f}^d, \mathbf{I}^D) P(\boldsymbol{\xi} | \mathbf{f}^d, \mathbf{I}^D)$. From (3.6), we see that

$$P(\mathbf{N} | \boldsymbol{\xi}, \mathbf{f}^d, \mathbf{I}^D) = \delta \left(\mathbf{N} - (\mathbf{G}^d \mathbf{f}^d + \boldsymbol{\xi}) \right). \quad (3.11)$$

Finally, carrying out the integration on $\boldsymbol{\xi}$ we obtain

$$P(\mathbf{N} | \mathbf{f}^d, \mathbf{I}^D) \propto \exp -\frac{1}{2} \left(\mathbf{N} - \mathbf{G}^d \mathbf{f}^d \right)^T \Sigma^{-1} \left(\mathbf{N} - \mathbf{G}^d \mathbf{f}^d \right), \quad (3.12)$$

where we have dropped the index ξ of Σ_ξ , so that now $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ represents the covariance matrix associate with $\boldsymbol{\xi}$'s distribution.

The distribution of (3.12), interpreted as a function of the quantities $\mathbf{f}^d$, may have different behaviors depending on the degeneracy of the linear system $\mathbf{G}^d \mathbf{f}^d = \mathbf{N}$. If it is determined, (3.12) can be put in the form of a normal multivariate distribution of mean $(\mathbf{G}^T \Sigma^{-1} \mathbf{G})^{-1} \mathbf{G}^T \Sigma^{-1} \mathbf{N}$ and variance matrix $\mathbf{G}^T \Sigma^{-1} \mathbf{G}$. However, if the linear system is degenerate and admits an infinite number of solutions, (3.12) cannot be normalized and has the shape of a normal "gutter" (or at least in two dimensions it does), as it does not

fall to zero at infinity. But we can still rotate it in space by taking the SVD of $\mathbf{G}$, as explained in section 2.3, and find that, along some linear combinations of the variables f_k^d , (3.12) still has a normal distribution behavior. More precisely, if $\mathbf{u}_i$ and $\mathbf{v}_i$ are the eigenvectors of $\mathbf{G}^T \mathbf{G}$ and $\mathbf{G} \mathbf{G}^T$, then

$$\mathbf{N} - \mathbf{G} \mathbf{f}^d = \mathbf{N} - \sum_{i=1}^M s_i \mathbf{u}_i \mathbf{v}_i^T \mathbf{f}^d = \sum_{i=1}^M (N_{i,u} - s_i f_{i,v}^d) \mathbf{u}_i, \quad (3.13)$$

where $N_{i,u}$ are the components of $\mathbf{N}$ in the $\{u_k\}$ basis, $f_{i,v}^d$ the components of $\mathbf{f}^d$ in the $\{v_k\}$ basis and s_i the singular values of $\mathbf{G}$. Note that no information is given about the other $|\mathbf{f}^d| - M$ components $f_{i,v}^d$ in (3.13); they are free to assume any value because there are no sufficient data available to constrain the solutions. However, although the linear system is undetermined, equation (3.13) shows that we can still extract some type of information from it in some particular directions.

This is particularly interesting for the multigrid approach because we can start the iterative method from some coarse scale at which the linear system $\mathbf{G}^d \mathbf{f}^d = \mathbf{N}$ is determined. Once the grid size becomes larger than M , we start extracting information from the likelihood distribution for some particular linear combinations of the components of $\mathbf{f}^d$, while others remain uninformative.

It is also important to note that (3.12) has the functional form it has only because the errors $\boldsymbol{\xi}$ were assumed to be normally distributed. If the errors were distributed by some other probability family, such as the Poisson distribution, the likelihood would also assume that new form. Although many parts of the analysis made so far makes explicit reference to functional form of (3.12), similar conclusions could be made for other types of distributions.

3.2.4 Priors

As the likelihoods in the previous section, we also have to deal with a set of prior distributions. The first one, at the coarser scale, serves as an input of the model and is used to obtain a posterior at the same scale. This posterior is then used to construct a prior for the next scale. So the priors at the finer scales are somewhat fixed by the iterative model.

These priors are used to assign probabilities to the different permitted values of f_k^d on cell k . This includes assigning zero probabilities to values which the background information definitely excludes. For example, if V represents a star and f_k^d measures the mass bounded in some region k , then we can say with certain that f_k^d is restricted to positive values smaller or equal to the total mass of that star.

But the priors also have another practical role, which is to regularize the likelihood. Being a probability distribution properly normalized, it always falls to zero at some point. So, although the likelihood may allow an infinite number of solutions, the prior, based on our previous knowledge, allows only a finite region of the possible values of $\mathbf{f}^d$ to be non-zero (or almost non-zero).

Which family of distributions to use depends on the specific nature of $f(\chi)$. But we can constrain the number of families by making some general simplifications. We will assume that each grid cell is represented independently by a univariate distribution $P(f_k^d | \mathbf{I}^D)$ and

the join prior distribution given by the product of the single priors:

$$P(\mathbf{f}^d|\mathbf{I}^D) = \prod_{k \in \Lambda_d} P(f_k^d|\mathbf{I}^D) , \quad (3.14)$$

which corresponds to a mean field approximation of $P(\mathbf{f}^d|\mathbf{I}^D)$.

3.3 Learning Dynamics

The process of learning, that is, the process of extracting information from the likelihood and conciliating it with the prior in order to generate a posterior is given, at any scale of resolution d , by Bayes' Theorem:

$$P(\mathbf{f}^d|\mathbf{N}, \mathbf{I}^D) = P(\mathbf{f}^d|\mathbf{I}^D) \frac{P(\mathbf{N}|\mathbf{f}^d, \mathbf{I}^D)}{P(\mathbf{N}|\mathbf{I}^D)} . \quad (3.15)$$

But suppose we start from the coarser scale $d = 0$ with a global prior $P(\mathbf{f}^0|\mathbf{I}^D) = \prod_{k \in \Lambda_0} P(f_k^0|\theta_k^0)$, a product of distributions over the sites of Λ_0 just like (3.14). The distributions $P(f_k^0|\theta_k^0)$ are taken to belong to a particular probability family $\mathcal{P}$ parametrized by θ_k^0 , collectively written as $\boldsymbol{\theta}^0$. The problem we face, once the prior is updated through (3.15) into $P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)$, is to decide what will represent the prior of the next scale. The main difficulty is that, by going from a scale to another, we multiply the dimension of $\mathbf{f}^0$ by the number of subdivisions each grid cell undergoes. For instance, if every cell of scale $d = 0$ is subdivided into K new cells, then $|\mathbf{f}^1| = K|\mathbf{f}^0|$. So we can't simply use the posterior of the previous scale as the prior of the next scale as they refer to a different number of variables. What we can do, however, is to relate parent and daughter grid cells. The idea is to make daughter cells inherit their prior distributions from their parent cells' posteriors. This is shown in figure 3.2.

Figure 3.2: Example of the inheritance scheme of posteriors.

In that approach, the simplest choice to make is

$$P(f_{k(j)}^1|\mathbf{I}^D) = P(f_j^0|\mathbf{N}, \mathbf{I}^D) , \quad (3.16)$$

where the index $k(j)$ indicates that the grid cell k is a daughter of j . That is, we simply use the posterior of the previous scale as the prior for the next scale. However, in order to do that, we must know what $P(f_j^0|\mathbf{N}, \mathbf{I}^D)$ is, which is not trivial as $P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)$ is not necessarily a product of distributions over the grid cells.

So far we have assumed that the likelihood has the functional form of (3.12). As we mentioned in the end of section 3.2.3, it need not be the case; it depends on the error

associated with $\mathbf{N}$. Let us call $\mathcal{L}$ the particular probability family to which the probability distribution of ξ belongs. In some cases, given that $\mathcal{L} \neq \mathcal{P}$, the posterior distribution will not necessarily belong to $\mathcal{P}$. This is the reason why we cannot assume $P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)$ to trivially give $P(f_j^0|\mathbf{N}, \mathbf{I}^D)$. What we have to do is to use a mean field approximation of the posterior $P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)$, as we have explained in section 2.2.3:

$$P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D) \rightarrow Q(\mathbf{f}^0|\mathbf{I}^D) = \prod_{k \in \Lambda_0} P(f_k^0|\zeta_k^0) . \quad (3.17)$$

Following the steps of section 2.2.3, we construct the entropy functional of $P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)$ relative to $Q(\mathbf{f}^0|\mathbf{I}^D)$,

$$S[P|Q] = - \int \prod_{k \in \Lambda_0} df_k^0 P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D) \log \frac{P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)}{\prod_{k \in \Lambda_0} P(f_k^0|\zeta_k^0)} , \quad (3.18)$$

and we maximize it in respect to ζ_k^0 to obtain, for every k ,

$$\left\langle \frac{\partial P(f_k^0|\zeta_k^0)/\partial \zeta_k^0}{P(f_k^0|\zeta_k^0)} \right\rangle_{P(\mathbf{f}^0|\mathbf{N}, \mathbf{I}^D)} = 0 . \quad (3.19)$$

Once a specific family of distribution is given, we can further develop the equations provided by (3.19) and obtain the parameters ζ_k^0 .

What (3.15) and (3.19) together amounts, in effect, is to a mapping

$$\theta_k^0 \rightarrow \zeta_k^0 , \quad k \in \Lambda_0 . \quad (3.20)$$

So now we can now give meaning to (3.16). Each grid cell k at $d = 0$ now has an associated posterior distribution that can be represented by ζ_k^0 . Going back to (3.16), we are free to define the prior distribution's parameters θ_k^1 as functions g_k^1 of the parents' cells parameters:

$$\theta_{j(k)}^1 = g_{j(k)}^1(\zeta_k^0) . \quad (3.21)$$

The simplest choice represented in (3.16) is now obtained by $g(x) = x$. But why should we prefer the identity over any other function? This is a rather crucial point. As we will see in chapter 4, different choices of g can lead to very distinct final results. One reason behind this is that g_k^1 controls the weight, or preference, we give to prior information in contraposition to the likelihood.

By going from a scale d to the next one $d + 1$, the number of variables increases and the linear system $\mathbf{N} = \mathbf{G}^d \mathbf{f}^d$ becomes less constrained. Consequently, the shape of the likelihood changes, it becomes "broader". The lesson here is that with more variables, the likelihood becomes less informative. Therefore, it is plausible that one may obtain a very thin posterior at $d = 0$, the coarser scale. So far so good. But then we would use this very thin distribution as the prior for $d = 1$ and it would represent a state of high certainty in comparison to the new likelihood. That means we would learn very little about $\mathbf{f}^d$, not only at $d = 1$ but also in all the subsequent iterations. One way to cope with this is to, in some way, renormalize the parameters ζ_k^0 using g .

3.3.1 Parameter Renormalization

So suppose we now have chosen the new set of parameters ζ^d that will describe the posterior $P(\mathbf{f}^d|\mathbf{N}, \mathbf{I}^D)$ by a mean field approximation $Q(\mathbf{f}^d|\mathbf{I}^D)$ according to the criteria imposed by (3.19). We proceed to refine our lattice Λ_d into Λ_{d+1} and we define new parameters $\theta_{j(k)}^{d+1} = g_{j(k)}^{d+1}(\zeta_k^d)$ which inherits the properties of ζ_k^d through some predefined function g . The problem we face now is to build some insight of what g should be.

Our main assumption is that the variables defined over a particular cell k are related to the variables of the daughter cells $j(k)$ by some function h :

$$f_k^d = h(f_{1(k)}^{d+1}, \dots, f_{K(k)}^{d+1}), \quad (3.22)$$

where K (in uppercase) represents the number of daughter cells k has. As an example, if f represents the mass or density of some substance, then we may use (3.22) as a way to assure mass or density conservation. In general,

$$f_k^d = C \sum_{j=1}^K f_{j(k)}^{d+1}. \quad (3.23)$$

We refer to relation (3.22) as $\mathcal{H}$. It states that if we, somehow, knew the variables $\{f_{j(k)}^{d+1}\}$, then f_k^d could be calculated right away. In terms of probabilities,

$$P(f_k^d | \{f_{j(k)}^{d+1}\}, \mathcal{H}) = \delta(f_k^d - h(f_{1(k)}^{d+1}, \dots, f_{K(k)}^{d+1})). \quad (3.24)$$

For now on, we will assume $\mathcal{H}$ is included in the set of background information $\mathbf{I}^D$.

In order to investigate this further, we first write the distribution of each f_k^d as a marginal over its generated cell parameters $\{f_{j(k)}^{d+1}\}$ and use the product rule given by standard probability theory to obtain

$$P(f_k^d | \mathbf{I}^D) = \int P(f_k^d | \{f_{j(k)}^{d+1}\}, \mathbf{I}^D) P(\{f_{j(k)}^{d+1}\} | \mathbf{I}^D) \prod_i df_{i(k)}^{d+1}. \quad (3.25)$$

Using (3.24),

$$P(f_k^d) = \int \delta(f_k^d - h(f_{1(k)}^{d+1}, \dots, f_{K(k)}^{d+1})) P(\{f_{i(k)}^{d+1}\} | \mathbf{I}^D) \prod_i df_{i(k)}^{d+1}. \quad (3.26)$$

Taking the Fourier transform of the distribution $P(f_k^d | \mathbf{I}^D)$ we obtain its characteristic function $\phi_k^d(\omega)$ and equation (3.26) becomes

$$\phi_k^d(\omega) = \int P(\{f_{i(k)}^{d+1}\} | \mathbf{I}^D) \exp \left\{ -i\omega h(f_{1(k)}^{d+1}, \dots, f_{K(k)}^{d+1}) \right\} \prod_i df_{i(k)}^{d+1}. \quad (3.27)$$

Now suppose we are allowed to approximate the joint probability $P(\{f_{i(k)}^{d+1}\} | \mathbf{I}^D)$ by a product of disjoint probability densities, just like we did with the priors in (3.14), and

assume h given by (3.23). We obtain

$$\phi_k^d(\omega) = \prod_i^K \phi_{i(k)}^{d+1}(C\omega) . \quad (3.28)$$

Then, one possible solution is to take

$$\phi_{i(k)}^{d+1}(\omega) = \phi_k^d\left(\frac{\omega}{C}\right)^{1/K} , \quad (3.29)$$

for each $i = 1, \dots, K$. The new parameters $\theta_{j(k)}^{d+1}$ can be chosen according to the solution of (3.29), depending on the specific family of prior distributions we are using.

The so obtained distribution parametrized by θ^{d+1} will now serve as the prior distribution to the next scale $d+1$. The entire process described so far can thus be resumed as a series of iterations of expressions (3.20) and (3.21):

$$\theta^0 \xrightarrow{(3.21)} \zeta^0 \xrightarrow{(3.20)} \theta^1 \xrightarrow{(3.21)} \dots \xrightarrow{(3.21)} \theta^d \xrightarrow{(3.20)} \zeta^D , \quad (3.30)$$

and it may be carried on until the maximum possible degree of resolution D is achieved.

3.4 Prior Examples

Since we are mainly interested in positive quantities in our applications to the geo-neutrino problem, we describe in details the procedure using gamma distributions as priors. As we explained earlier, each grid cell will be represented by a gamma distribution and the global prior will be given by their product. However, the method is not restricted to gammas and other distributions may as well be used.

Note that this is also input information. We are choosing by hand the functional form of the prior distribution instead of assuming we have no information before data is collected. That is, before incorporating the data, we are assuming we at least know what the variables $\mathbf{f}$ represent and how it is distributed.

3.4.1 Gamma Distributions

If we are dealing with positive-definite quantities $\mathbf{f}^d$, one possible choice to describe the probability distribution of each f_k^d is the gamma distribution, given by

$$\Gamma(f|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} f^{\alpha-1} \exp -\beta f , \quad (3.31)$$

with mean $\langle f \rangle = \alpha/\beta$ and variance $\text{var}(f) = \alpha/\beta^2$. For negative values of f it is defined as zero. In the same footing of section 2.2.2, this distribution is obtained when one imposes normalization, a mean value and a mean value of the logarithm.

The online updating step described in (3.19) can be easily calculated in this case since

$$\frac{1}{\Gamma} \frac{\partial \Gamma}{\partial \alpha} = \frac{\alpha}{\beta} - f , \quad (3.32)$$

$$\frac{1}{\Gamma} \frac{\partial \Gamma}{\partial \beta} = \log \beta - \Psi_1(\alpha) + \log f , \quad (3.33)$$

where Ψ_1 is the Digamma function. We obtain two sets of equations by carrying out the averaging over the posterior distribution at some scale d :

$$\frac{\alpha_k^d}{\beta_k^d} - \langle f_k^d \rangle = 0 , \quad (3.34)$$

$$\log \beta_k^d - \Psi_1(\alpha_k^d) + \langle \log f_k^d \rangle = 0 . \quad (3.35)$$

Together, (3.34) and (3.35) yield

$$\log \alpha_k^d - \Psi_1(\alpha_k^d) - \log \langle f_k^d \rangle + \langle \log f_k^d \rangle = 0 , \quad (3.36)$$

which may easily be solved numerically for α_k^d and (3.34) subsequently solves for β_k^d .

Once we have calculated the new set of parameters $\zeta^d = (\alpha^d, \beta^d)$ we can proceed to the next scale $d + 1$ by subdividing each cell k of scale d . We then may use the steps of section 3.3.1 to determine the prior distribution of scale $d + 1$. Since the characteristic function of the gamma distribution is given by

$$\phi(\omega) = \left(1 - i \frac{\omega}{\beta}\right)^{-\alpha} , \quad (3.37)$$

equation (3.29) gives $\alpha^{d+1} = \alpha^d/K$ and $\beta^{d+1} = \beta^d/C$. In the case when relation 3.23 happens to be a block average over the daughter grid cells, C becomes equal to K and we can write simply $\zeta^{d+1} = \zeta^d/K$.

3.4.2 Reducing Dimensions

As stated in equation (3.14), the joint prior distribution is obtained by taking the product of the priors of every grid cell. But we need not to necessarily carry all the cells along the whole iterative process, we may choose to exclude some of them when passing from one scale to another. A reason for this would be a previously imposed cutoff scale for the minimum amount of mass a cell can accommodate. For example, suppose we happen to know that the distribution of $\mathbf{f}$ inside V consists of a single, continuous and small region $F \subseteq V$ of non-zero density. Then, if after iteration d we are able to infer that some subregion of $U \subseteq V$ is highly unlikely to carry any content of $\mathbf{f}$, then we may exclude it. In practice, we remove grid cells lying inside U by excluding their corresponding variables from both the prior and the likelihood.

4 | Applications

We first apply and discuss the techniques described in the previous chapters to a very simple and easy to visualize bidimensional model. Then we investigate the 3D model.

4.1 General Procedure

Our applications, following the geo-neutrino problem, will consist of reconstructing positive charge distributions ρ , located inside some region V , from a set of data $\mathbf{N}$ collected by a number of detectors. We denote the surface of V by ∂V and we assume that the detectors all lie at ∂V . Note that are not necessarily making reference to electronic charges or masses, but to any kind of positive quantity that can be associated to the measured data via a relation analogous to (3.1).

This quantity N is measured by each detector a number of times, resulting in a set of normal distributions $\mathcal{N}(N_i, \sigma_i)$ centered at N_i and spread by σ_i . We take the means N_i to represent the data given by each detector i and to be used, along with σ_i , as input to the model. As explained in section 3.2.1, discretization at the zeroth order yields

$$\mathbf{N} = \mathbf{G}\boldsymbol{\rho} + \boldsymbol{\xi} , \quad (4.1)$$

where $\mathbf{G}$ is the theory's Green function, $\boldsymbol{\rho}$ is the charge distribution vector and $\boldsymbol{\xi}$ the experimental error distributed by $\mathcal{N}(0, \boldsymbol{\sigma}) = \prod_i \mathcal{N}(0, \sigma_i)$. At any scale of resolution, the likelihood distribution linking $\boldsymbol{\rho}$ to $\mathbf{N}$ is

$$P(\mathbf{N}|\boldsymbol{\rho}, \mathbf{I}) \propto \exp -\frac{1}{2}(\mathbf{N} - \mathbf{G}\boldsymbol{\rho})^T \boldsymbol{\Sigma}^{-1}(\mathbf{N} - \mathbf{G}\boldsymbol{\rho}) , \quad (4.2)$$

where we have omitted the normalization constant, $\mathbf{I}$ represents all the background information and $\boldsymbol{\Sigma}$, as explained in section 3.2.2, may involve both the experimental and the discretization errors. The prior $P(\boldsymbol{\rho}|\mathbf{I})$ is assumed to be given and we will denote its family of distributions by $\mathcal{P}$.

Algorithm

We start the iterative process by initializing the likelihood $P(\mathbf{N}|\boldsymbol{\rho}^0, \mathbf{I}^0)$ at the coarser scale $d = 0$. Because we are interested in spacial correlations between the grid cells and the detectors, initializing $P(\mathbf{N}|\boldsymbol{\rho}^0, \mathbf{I}^0)$ generally involves calculating the distances between every grid cell and detector and storing some function of the calculated values in a matrix $\mathbf{G}^0$, as shown in algorithm 1.

Algorithm 1 Setup the Green Function Matrix

```

1: procedure SETUP_GREEN( $d$ )
2:    $\mathbf{X}_N \leftarrow$  detector's positions
3:    $\mathbf{X}_{G^d} \leftarrow$  grid cell's positions ▷ Depends on the chosen discretization
4:    $d\mathbf{v} \leftarrow$  volumes of grid cells
5:    $\mathbf{G}_{NG^d} \leftarrow \mathbf{G}(d\mathbf{v}, |\mathbf{X}_N - \mathbf{X}_{G^d}|)$  ▷  $\mathbf{G}$  is the Green function of the specific problem
6:   return  $\mathbf{G}_{NG^d}$ 

```

Next we calculate the non-normalized posterior distribution $P(\boldsymbol{\rho}^0 | \boldsymbol{\phi}, \mathbf{I}^0)$ using Bayes' Theorem and run a Metropolis algorithm (see algorithm box 4) to calculate any function F of $\boldsymbol{\rho}^0$ averaged over the posterior equation (3.19) requires us to obtain. In the case of gamma distributions, as explained in section 3.4.1, we would have to calculate $\langle \boldsymbol{\rho}^0 \rangle$ and $\langle \log \boldsymbol{\rho}^0 \rangle$. So we calculate $\langle F(\boldsymbol{\rho}^0) \rangle$:

$$\langle F(\boldsymbol{\rho}^0) \rangle = \frac{1}{A} \int \prod_i d\rho_i^0 F(\boldsymbol{\rho}^0) P(\boldsymbol{\rho}^0 | \mathbf{I}^0) \exp -\frac{1}{2}(\mathbf{N} - \mathbf{G}^0 \boldsymbol{\rho}^0)^T \boldsymbol{\Sigma}^{-1} (\mathbf{N} - \mathbf{G}^0 \boldsymbol{\rho}^0), \quad (4.3)$$

where A is a normalization constant. This a necessary step in order to calculate the quantities that appear in (3.19). Note, as shown in algorithm 4, that in order to avoid precision problems, we use the logarithm of the Metropolis' weight w . We also simplify the integrand of (4.3) to speed up computations and assume a non-symmetric proposal distribution $Q(\boldsymbol{\rho}_n | \boldsymbol{\rho})$ to deal with the positive nature of $\boldsymbol{\rho}$. Thus, given a state $\boldsymbol{\rho}$, a proposed $\boldsymbol{\rho}_n$ is accepted with probability given by

$$\log w = \log \frac{P(\boldsymbol{\rho}_n | \mathbf{N}, \mathbf{I})}{P(\boldsymbol{\rho} | \mathbf{N}, \mathbf{I})} \frac{Q(\boldsymbol{\rho} | \boldsymbol{\rho}_n)}{Q(\boldsymbol{\rho}_n | \boldsymbol{\rho})} = \log w_0 + \log w_L + \log w_Q, \quad (4.4)$$

where w_0 represents the weight due to the prior, w_L the weight due to the likelihood and w_Q the weight due to the proposal distribution. If we further assume that each component of $\boldsymbol{\rho}$ is atualized at a time, that is, $\rho_i \rightarrow (\rho_i)_n = \rho_i + \delta\rho\delta_{ik}$, then w_L is simplified to

$$\begin{aligned} \log w_L &= \frac{1}{2} \left\{ (\mathbf{N} - \mathbf{G}\boldsymbol{\rho})^T \boldsymbol{\Sigma}^{-1} (\mathbf{N} - \mathbf{G}\boldsymbol{\rho}) - (\mathbf{N} - \mathbf{G}\boldsymbol{\rho}_n)^T \boldsymbol{\Sigma}^{-1} (\mathbf{N} - \mathbf{G}\boldsymbol{\rho}_n) \right\} \\ &= (\mathbf{N}^T \boldsymbol{\Sigma}^{-1} \mathbf{G})_k \delta\rho + \frac{1}{2} (\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G})_{kk} (\delta\rho)^2 - (\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\boldsymbol{\rho})_k \delta\rho, \end{aligned} \quad (4.5)$$

greatly decreasing the algorithm's runtime.

Once all the $\langle F(\boldsymbol{\rho}^0) \rangle$ are calculated, we use them to solve equations (3.19) via Newton's method (see algorithm box 3, where we have exemplified it for the case of gamma distributions). Then, we refine the grid to the next scale $d = 1$ by dividing each cell into a fixed number of new cells and inheriting to scale $d = 1$ the parameters of scale $d = 0$. Finally, we renormalize the new parameters according to what we discussed in section 3.3.1 and proceed to repeat the process on the next finer scale.

The process is iterated until we perceive we can no longer extract new information from the data. Algorithm 2 summarizes the entire process in algorithmic form.

Once we have obtained the finest resolution D grid, we expect to be able to infer the

Algorithm 2 Iterative Process

```

1: procedure ITERATE( $\theta^0, \mathbf{N}, D$ )
2:   for  $d = 0$  to  $D$  do
3:      $\mathbf{G}^d \leftarrow \text{setup\_green}(d)$  ▷ Setup the Green matrix at resolution  $d$ 
4:      $\langle F(\rho^d) \rangle \leftarrow \text{metropolis}(d, \theta^d, \mathbf{N}, \sigma^d, \mathbf{G}^d)$  ▷ Metropolis' algorithm
5:      $\theta^d \leftarrow \text{newton}(d, \langle F(\rho^d) \rangle)$  ▷ Newton-Raphson method
6:      $\theta^{d+1} \leftarrow \text{inherit}(\theta^d)$  ▷ Setup the parameters for the next finer scale  $d + 1$ 
7:      $\theta^d \leftarrow \text{renormalize}(\theta^d)$  ▷ Renormalize the parameters according to sec. 3.3.1

```

Algorithm 3 Newton's Method for Γ distributions

```

1: procedure NEWTON( $d, \langle \rho^d \rangle, \langle \log \rho^d \rangle$ )
2:    $\theta \leftarrow \theta_0$  ▷ Set initial conditions for  $\theta$ 
3:   for  $i = 0$  to  $N$  do
4:      $\mathbf{f}_\theta \leftarrow \log \theta - \Psi_0(\theta) + \langle \log \rho^d \rangle - \log \langle \rho^d \rangle$ 
5:      $d\mathbf{f}_\theta \leftarrow 1/\theta - \Psi_1(\theta)$ 
6:      $\theta \leftarrow \theta - \mathbf{f}_\theta / d\mathbf{f}_\theta$ 
7:   return  $\theta$ 

```

Algorithm 4 Metropolis' Algorithm

```

1: procedure METROPOLIS( $d, \theta^d, \mathbf{N}, \mathbf{G}^d, \Sigma$ )
2:   Set number of samples  $S$ , step size step and burn-in period burn
3:    $\mathbf{y} \leftarrow \mathbf{y}_0$  ▷ Set initial position of the Markov chain
4:   for  $i = 0$  to  $S$  do
5:     for  $j = 0$  to  $|\rho^d|$  do
6:        $dy \leftarrow \text{random}(-\text{step}, \text{step})$ 
7:        $y_{\text{new}} \leftarrow y_j + dy$ 
8:       while  $y_{\text{new}} \leq 0$  do
9:          $dy \leftarrow \text{random}(-\text{step}, \text{step})$ 
10:         $y_{\text{new}} \leftarrow y_j + dy$ 
11:         $w_L \leftarrow (\mathbf{N}^T \Sigma^{-1} \mathbf{G})_j dy + \frac{1}{2} (\mathbf{G}^T \Sigma^{-1} \mathbf{G})_{jj} (dy)^2 - (\mathbf{G}^T \Sigma^{-1} \mathbf{G} \mathbf{y})_j dy$ 
12:         $w_0 \leftarrow \log P(\mathbf{y}_{\text{new}} | \mathbf{I}^d) / P(\mathbf{y} | \mathbf{I}^d)$ 
13:         $w_Q \leftarrow \log Q(\mathbf{y} | \mathbf{y}_{\text{new}}) / Q(\mathbf{y}_{\text{new}} | \mathbf{y})$ 
14:         $w \leftarrow w_0 + w_L + w_Q$ 
15:        if  $w > 0$  or  $w > \log \text{random}(0, 1)$  then ▷ Metropolis' acceptance weight
16:           $y_j \leftarrow y_{\text{new}}$ 
17:          if  $i > \text{burn}$  then ▷ Collect the generated samples
18:             $\langle F(\mathbf{y}) \rangle \leftarrow \langle F(\mathbf{y}) \rangle + F(\mathbf{y})$ 
19:             $\text{samples} \leftarrow \text{samples} + 1$ 
20:   return  $\langle F(\mathbf{y}) \rangle / \text{samples}$ 

```

total content of positive charge ρ inside of V by summing every ρ_k^D . In the case where ρ represents a density, the total charge is calculated by summing each ρ_k^D multiplied by the respective cell volume δv_k^D .

4.2 Toy Model in Two Dimensions

To test the method in a very simple way, we propose a toy model in two dimensions. We could investigate directly the ideas of chapter 3 in a 3-dimensional model, exemplifying the application to geo-neutrinos. However, there are a few tests and remarks we would like to make first and not only the bidimensional model is easier to visualize but it presents the same behaviors as we observe in the 3D case.

We place a positive charge distribution ρ in a region bounded by a circle of unit radius and M detectors on the circle. Each detector then provides a distribution $\mathcal{N}(N_i, \sigma_i)$ representing its measurements, where each σ_i is chosen by hand. To gain some physical insight, we may interpret ρ as the mass density of some radioactive element and the measured quantities N_i the number of particles detector i measures over a period of time. We may also multiply each ρ_k^d by the volumes δv_k^d of each cell and interpret $m_k^d = \delta v_k^d \rho_k^d$ as the total mass bounded inside cell k . As we are in two dimensions, we decide to use $\mathbf{G}(\mathbf{R}, \mathbf{r}) = 1/2\pi|\mathbf{R} - \mathbf{r}|$.

The prior at the coarser scale is assumed to be a product of gamma distributions, each representing the possible amount of mass in the cells of the initial grid. Because initially we have no information whatsoever about the amount of mass in each cell, we adopt the the most ignorant prior we can, which amounts to choosing constant parameters $\alpha_i^0 = \alpha^0$ and $\beta_i^0 = \beta^0$ between the grid cells.

As we have discussed in section 2.1.2, there is a certain degree of arbitrariness in deciding what number we will extract from the posterior distributions to represent the mass density or the total amount of mass of each grid cell. As in (2.10), we will resort to the posterior mean along with some measure of the variance. To help visualization we will also show the distributions themselves.

4.2.1 Experimental Setup

The initial grid is chosen to be a 4 cell approximation to the unity circle. At the end of each iteration, each of them divides itself into 4 new cells and we are left with a grid of size 4^{d+1} , where d is the degree of resolution index. We investigate a mass distribution located near to the first coarser grid and normalized to 1. Figures 4.1 show the particular grid scheme constructed in a series of iterations, the particular chosen constant mass distribution inside the circle and the detectors chosen at random on the surface.

Here, the total number of detectors was chosen to be $M = 30$, each having a random accuracy of about 1%. That is, each detector was assumed to have a σ near 0.01 times its measurement N_i .

It is important to note that the simulations below were carried in the density variables ρ . So, effectively, the Green function contains a cell volume factor δv . In order to interpret the plots below in terms of total mass, one must multiply the density values by the factor δv , given simply by the total volume of the unit circle divide by the number of cells 4^d .

Figure 4.1: Constructed grids at the initial, second and third scales. They are 4^{d+1} cell approximations to the unity circle, represented by points at locations indicated by black dots. The mass distribution is shown in a black strip and the detectors' locations are represented by spikes on surface of the circle. Note that in this particular construction, the cells all have the same area.

4.2.2 Results and Discussion

We investigate some particular situations in the presence or absence of some of the tools we have discussed in chapter 3.

First, suppose we set the discretization error of section 3.2.1 to zero, $\eta = 0$, and we also do not renormalize the parameters when going from one scale to another; we simply pass $\zeta^d = (\alpha^d, \beta^d)$ to the cells of scale $d + 1$. The only variables we have left to modify are the parameters of the prior of the initial grid, (α^0, β^0) . Taking $\alpha_k^0 = 1$ and $\beta_k^0 = 1$ (we choose high variance distributions to represent our lack of knowledge of the amount of mass in each cell of the coarser grid before data is implemented), we obtain the results shown in figures 4.2, 4.3, 4.4 and 4.21. We see that we are unable to reconstruct the original distribution. The total amount of mass, obtained by summing all the mean values multiplied by the volume factor δv_k^d , is also poorly reconstructed, amounting for 0.8991190 ± 0.0000002 , whereas we expected to obtain 1.0.

Figure 4.2: Results after the first scale iteration.

Figure 4.3: Results after the second scale iteration.

Figure 4.4: Results after the third scale iteration.

Figure 4.5: Results after the fourth scale iteration.

Do these reconstructions at least do justice to the experimental information $\mathbf{N}$ we have? That is, do they really generate the data used as input? To check this, we use the mean values shown in the figures above to represent the amount of mass in each cell and the variances to represent our uncertainties. We then calculate the N_i that these reconstructions generate at each detector. From figures 4.6 we see that we are also unable to match the experimental measurements, at least to the extent of the errors (the error bars are smaller than the points).

Figure 4.6: Direct and Inverse experimental measurements by the detectors at each scale. The inverse measurements are calculated assuming the mean values of the distributions to represent the amount of mass in each grid cell.

On the other hand, by taking advantage of the large number of detectors, we can also start the process at the second grid. Suppose we do this, so that the coarser one is now $d = 2$. The results are shown in figures 4.7, 4.8, 4.9 and 4.10. By comparing figures 4.9 and 4.21, we see that 4.9 looks like a better result. However, we still have discrepancies from the direct and inverse measurements (see figure 4.10) and the total amount of mass 0.96121 ± 0.00007 remains poorly determined. This is an indication that the process of passing from one scale to another is not behaving well.

Figure 4.7: Results after the first scale iteration.

Figure 4.8: Results after the second scale iteration.

Figure 4.9: Results after the third scale iteration.

Figure 4.10: Direct and Inverse experimental measurements by the detectors at each scale, corresponding to the second example.

In order to quantify how accurate these reconstructions are, we define a normalized vector $\mathbf{m}^D$ at the finer scale $d = D$. $\mathbf{m}^D$ takes the input mass distribution (the one we used in the direct problem to calculate the experimental measurements N_i) and uses it to assign to its components the amount of mass lying inside of each cell. So, for instance, in the example of figure 4.1 we would have $D = 2$ and $\mathbf{m}^D$ would be something like $(0, \dots, 0, m_{k_1}, m_{k_2}, 0, \dots, 0)$, since the charge distribution lies entirely inside only two cells of grid D . We then define at each scale $d = 1, \dots, D$ normalized vectors $\mathbf{m}_0^d$ of the same dimension of $\mathbf{m}^D$ that take the inferred mass amounts in each cell after iteration d and throws them inside their daughter grid cells of scale $d = D$. The dot product $\cos \beta = \mathbf{m}^D \cdot \mathbf{m}_0^d$ measures how well our inference approximates to the best result we could expect to obtain. Figure 4.11 shows $\mathbf{m}^D \cdot \mathbf{m}_0^d$ for both the examples given above. We see that results for the second example are more accurate.

(a) Results for the first example.

(b) Results for the second example.

Figure 4.11: $\mathbf{m}^D \cdot \mathbf{m}_0^d$ for the two first examples given. Each point corresponds to an iteration.

Let us now turn on the discretization error and the renormalization factor. Following section 3.4.1, at the end of each iteration we divide α and β by the number of daughter cells, $K = 4$. As proposed in section 3.2.2, we take the discretization error to be $\gamma 4^{-d}$, γ being a proportionality factor and 4^d the grid size. For $\gamma = 2$ we obtain the results shown in figures 4.12, 4.13, 4.14, 4.15, 4.16 and 4.17. The reconstruction, specially at the third iteration (figure 4.14), seems to approximate much better the real distribution. The total inferred mass is now 0.92 ± 0.14 .

There is a drawback, however, in adding these discretization uncertainties. The distributions are now much more spread along the x axis. That is, their variance are much larger now. Consequently, the choice of a point in each distribution to represent the amount of mass $\mathbf{m}^d$ lying inside the cells becomes much more subjective than simply choosing the mean value. This subjectiveness lead to large errors in any quantity we calculate using $\mathbf{m}^d$, such as the inferred data in figure 4.16 and the dot products $\mathbf{m}^D \cdot \mathbf{m}_0^d$ in figure 4.17.

(a) Mean values of the distributions in each cell.

(b) Variance of the distributions in each cell.

(c) Gamma distributions in each grid cell. The colors match the mean values on the left.

Figure 4.12: Results after the first scale iteration.

Figure 4.13: Results after the second scale iteration.

Figure 4.14: Results after the third scale iteration.

Figure 4.15: Results after the fourth scale iteration.

Figure 4.16: Direct and Inverse experimental measurements by the detectors at each scale, corresponding to the third example.

Figure 4.17: $\mathbf{m}^D \cdot \mathbf{m}_0^d$ for the third examples given. Each point corresponds to an iteration.

What went wrong with the two first examples in which we did not take into account the corrections discussed in chapter 3? First we started at a scale where we had only 4 slots to decide where the whole mass content should be placed at, with a complete ignorant prior. Since each detector had a significantly high measurement precision, the result reflected a high certainty of the amount of mass inside each of the 4 cells, and our initial prior did nothing to oppose this change in beliefs. Nevertheless, this is a very plausible situation. We then tried to use these distributions as priors to the next scale, but the new likelihood distribution could barely shift them. The same occurred in the second example, although in that case we had 16 initial slots to make the inference. If we plot the prior and posterior family chain of a grid cell belonging to the finer scale, that is, all the priors and posteriors that led to a specific posterior in the last grid, we can see the behavior of the learning dynamics. Figures 4.18 and 4.19 show this for the first and second examples, respectively.

Figure 4.18: Dynamics of learning from a prior (in grey) to a posterior (in red) at each scale of the first example shown. Note that the prior of a scale is the posterior of the previous scale with the variance 4 times larger.

Figure 4.19: Dynamics of learning for the second example.

On the other hand, in the third example, where we have accounted for the corrections, the priors at each iteration were more apt to dislocate due to the likelihood. However, the variance increases drastically to the point of allowing us to draw almost no conclusions. Figure 4.20 shows a prior family chain in the third example.

Figure 4.20: Dynamics of learning for the third example.

Ideally, we would start from a coarse scale, obtain a posterior distribution, increase its variance by the exact necessary amount so that it becomes vulnerable (but not too much) to the likelihood of the next scale and pass it as a prior. We can try to reproduce this situation by instead of dividing the posterior parameters at the end of each iteration by 4, as we have been doing so far, dividing them by an arbitrary value. Figures 4.21 show the results at the last scale for this procedure, which turn out to be very similar to the results of the third example, except possibly that we might have superestimated the errors.

Figure 4.21: Results at the finest scale for the fourth example.

Thus, we will resort to the third examples' method in our applications to the 3-dimensional model.

4.3 Three Dimensional Model

Now we apply the techniques we used in the previous section to investigate a three dimensional version of the model. In the same way, we consider a sphere of unit radius and place M detectors on its surface. However, as we are now in three dimensions, we use $\mathbf{G}(\mathbf{R}, \mathbf{r}) = 1/4\pi|\mathbf{R} - \mathbf{r}|^2$. Again, the prior at the coarser scale is taken to be uniform among the grid cells, with $\alpha_k^0 = \alpha^0 = 1$ and $\beta_k^0 = \beta^0 = 1$.

4.3.1 Setup

The first approximation to the unit sphere is a 8 cell grid, with 4 in each hemisphere. After each iteration, the cells divide themselves into 8 new ones of equal volumes. So now the grid size is 8^d , where d is the resolution index, and the discretization error is $\gamma 8^{-2d/3}$.

Two different setups were constructed. The first one, shown in figure 4.22, is a single constant mass distribution located inside the region $[0.7, 0.8] \times [0.6\pi, 0.8\pi] \times [0.3\pi, 0.4\pi]$ (in spherical coordinates (r, ϕ, θ)). The second one, shown in figure 4.23, involves two distant regions of constant mass distributions, $[0.7, 0.8] \times [0.6\pi, 0.8\pi] \times [0.3\pi, 0.4\pi]$ and also $[0.6, 0.7] \times [1.5\pi, 1.8\pi] \times [0.5\pi, 0.7\pi]$, each containing 1 unit of mass. In both cases we have used $M = 150$ detectors.

Figure 4.22: First setup of the three dimensional model. The red dots represent the detectors distributed randomly at the surface of the unity sphere and the black dots the uniform distribution of mass.

Figure 4.23: Second setup of the three dimensional model. The black dots represent the two constant mass distributions placed inside the unity sphere. Note that their radial component are different.

Again, in both cases the errors σ_i were chosen to be random values near 0.01 times their corresponding measurements N_i .

4.3.2 Results and Discussion

The mean values and the variances for the first setup are shown in figures 4.26, 4.27 and 4.28, where we have projected each inner sphere into the plane. Figure 4.25 shows the gamma distributions corresponding to the grid cells after each iteration. We see that we are able to reconstruct the original distribution at least up to the third scale. However, again, the further we proceed to finer scales, the less constrained becomes the inferred values of mass inside each cell as the variances become larger. For this first example the total amount of inferred mass was 0.97 ± 0.04 .

(a) Direct measurements N_i obtained by the detectors. The white strip represents the projection of the original mass distribution.

(b) Inferred N_i using the third scale reconstruction of figure 4.28.

Figure 4.24: Direct and Inverse experimental measurements. To construct the surface we have interpolated the points.

Figure 4.25: Gamma distributions corresponding to the the mean values and variances shown in figures 4.26, 4.27 and 4.28, respectively. The colors match the mean values.

Figure 4.26: First scale results.

Figure 4.27: Second scale results.

Figure 4.28: Third scale results.

The results for the second setup with two separated mass distributions are shown in figures 4.29, 4.31, 4.32, 4.33 and 4.30. The total amount of inferred mass was 1.87 ± 0.11 .

(a) Direct measurements N_i obtained by the detectors. The white strips represent the projection of the original mass distributions.

(b) Inferred N_i using the third scale reconstruction of figure 4.33.

Figure 4.29: Direct and Inverse experimental measurements. To construct the surface we have interpolated the points.

(a) First scale.

(b) Second scale.

(c) Third scale.

Figure 4.30: Gamma distributions corresponding to the the mean values and variances shown in figures 4.31, 4.32 and 4.33, respectively. The colors match the mean values.

Figure 4.31: First scale results.

Figure 4.32: Second scale results.

Figure 4.33: Third scale results.

It is important to point out that the distributions used so far do not exhibit spherical symmetries. Spherically symmetric distributions produce the same N_i at every point on the surface of the sphere, making it impossible to access the bulk distribution. The same applies for the bidimensional case.

4.4 Geo-Neutrinos?

Whilst we are not able at the current geo-neutrino experimental state to reconstruct Earth's bulk content of ^{232}Th and ^{238}U , we can at least point out how to carry the method for future applications. First of all, the geo-neutrino number (1.3) can be written exactly in the form of (3.1) by defining

$$\mathbf{G}(\mathbf{R}_i, \mathbf{r}, X) = n_i^p t_i \langle \epsilon_i \rangle \langle \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \rangle \langle \sigma_i \rangle_X A_X \frac{\delta v_{\mathbf{r}}}{4\pi |\mathbf{R}_i - \mathbf{r}|^2} . \quad (4.6)$$

For ^{232}Th and ^{238}U , if we consider only current experiments that work with inverse beta decay, we obtain:

$$\mathbf{G}(\mathbf{R}_i, \mathbf{r}, ^{232}\text{Th}) = 3.43 \times 10^{-15} \langle \epsilon_i \rangle \left(\frac{n_i^p}{10^{32}} \right) \left(\frac{t_i}{\text{days}} \right) \frac{\delta v_{\mathbf{r}}}{|\mathbf{R}_i - \mathbf{r}|^2} \text{ km}^3/\text{kg} , \quad (4.7)$$

$$\mathbf{G}(\mathbf{R}_i, \mathbf{r}, ^{238}\text{U}) = 3.34 \times 10^{-14} \langle \epsilon_i \rangle \left(\frac{n_i^p}{10^{32}} \right) \left(\frac{t_i}{\text{days}} \right) \frac{\delta v_{\mathbf{r}}}{|\mathbf{R}_i - \mathbf{r}|^2} \text{ km}^3/\text{kg} , \quad (4.8)$$

where we have used $\langle \sigma \rangle_{\text{U}} = 0.404 \times 10^{-44} \text{ cm}^2$, $\langle \sigma \rangle_{\text{Th}} = 0.127 \times 10^{-44} \text{ cm}^2$ [Fiorentini et al., 2007], $\langle \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \rangle \simeq 0.58$ (see appendix A), $A_{\text{U}} = 12443 \text{ Bq/g}$ and $A_{\text{Th}} = 4059 \text{ Bq/g}$. Next, we use prior information to construct the grids and the prior distribution at the coarser scale. This scale may be taken as the finest one we can construct in the absence of experimental geo-neutrino data, but in the presence of other geophysical and geochemical information. For example, it has been asserted several times in the literature that, due to geochemical arguments, the Earth core's content of radiogenic material must be null [McDonough, 2003]. We may then, as a first approach, try to reconstruct the Earth's radiogenic content by assigning to the core's region grid cells with prior distributions lying entirely near the origin, or simply by excluding them. Next, we follow the steps of algorithm 2.

4.4.1 New Detectors

After we have obtained the finest possible degree of resolution from our data, we may then ask how should new data be gathered. In other words, where should we place the next detector in order to maximize the amount of information we can extract from data?

In the multigrid method, at each degree of resolution we have a posterior distribution that depends on the results of the previous scale. So, when new data Y is obtained, in order to calculate the information gain, we must calculate the mutual information 2.18 between the two different setups at the same degree of resolution d : the one which evolved to d with Y and the other which evolved without Y .

The idea is to place a new detector on the surface of the sphere at coordinates labeled by

ϵ and then calculate the posterior distribution $P(X|Y, I_\epsilon, Y_0)$ at some predefined maximum degree of resolution d . Next we calculate the entropy between $P(X|Y, I_\epsilon, Y_0)$ and $P(X|Y_0)$, the previously obtained posterior. As we do not know a priori Y , we take the mean of the entropy over $P(Y|Y_0, I_\epsilon)$. This amounts to the mutual information between $P(X|Y, I_\epsilon, Y_0)$ and $P(X|Y_0)$, the parameter ϵ being free to vary. To simplify calculations, we write 2.18 in the form

$$\begin{aligned} & \int dY dX P(X, Y|Y_0, I_\epsilon) \log \frac{P(X, Y|Y_0, I_\epsilon)}{P(Y|Y_0, I_\epsilon)P(X|Y_0)} \\ &= \int dY dX P(X, Y|Y_0, I_\epsilon) \log P(Y|X, Y_0, I_\epsilon) + \int dX P(X|Y_0, I_\epsilon) \log \frac{P(X|Y_0, I_\epsilon)}{P(X|Y_0)} \\ & \quad - \int dY P(Y|Y_0, I_\epsilon) \log P(Y|Y_0, I_\epsilon) . \end{aligned} \quad (4.9)$$

Note, however, that the first term in the second line of (4.9) involves the entropy of the likelihood $P(Y|X, Y_0, I_\epsilon)$ in relation to some constant prior. Now, the error ξ_i of each detector should not depend on ϵ ; if the experimental apparatus is the same, the error should be the same independently if the experiment is located in Brazil or in Japan. Consequently, the entropy of $P(Y|X, Y_0, I_\epsilon)$ must be independent of ϵ [Loredo, 2003]. Furthermore, if we had not included Y in our second run of the algorithm, we would have obtained the same answer as in our first run. Therefore, the second term in the second line of (4.9) must be zero. We are left with the task of maximizing the quantity

$$C - \int dY P(Y|Y_0, I_\epsilon) \log P(Y|Y_0, I_\epsilon) , \quad (4.10)$$

in relation to the coordinates ϵ ($\epsilon = (x, y, z)$ for example), where C is a constant representing the entropy of the likelihood. $P(Y|Y_0, I_\epsilon)$ can be obtained with the help of (2.19). However, we do not proceed with applications of this method in this dissertation, we leave it for a future work.

5 | Conclusions and Future Work

One of the main advantages of the treatment we have given to the geo-neutrino inverse problem is the possibility of easily incorporating different kinds of prior assumptions. We can modify the initial prior and the sequence of grids in a flexible way to include information from the Earth given by sources that do not concern directly the flux of geo-neutrinos but that pose bounds on it. We can thus explore both model independent measurements and measurements in the presence of other assumptions all in the same framework by just running the same algorithm. As an example given in section 4.4, we could include or not Earth's core in our simulations.

Currently, the drawback is that we still have very few detectors to draw geo-neutrino data from and the available data carry large uncertainties. So, taking real predictions from geo-neutrino data about the shape of the bulk distribution of radiogenic sources inside the Earth is still a matter for the future. What we can do now, however, is to progress in the subject by proposing new experiments. One way of doing this is by simply constructing new detectors anywhere on the Earth. However, we can now take advantage from the already obtained geo-neutrino data and see whether it contain information of where we should place the next detectors, or at least draw some preferred regions. This can be done by exploring the method of maximum entropy, as explained in section 2.2.4, in conjunction to the multigrid method.

The main difficult we have encountered in this multigrid approach was to build a practical theory that could tell us how to go from a scale to the next. It is clear from the first two examples of the model in two dimensions (see section 4.2.2) that some form of parameter renormalization or discretization error is needed in order to account for the small variance distributions. In our applications we have introduced a parameter γ that represented the intensity of the discretization errors and which was free for us to adjust. We observed that for large values of γ , we could simply not make any kind of precise prediction since the generated errors were gigantic. So, in general, a large γ must be associated with error superestimation. On the other hand, with a small γ , we would fall again in the case of distributions of extremely small variances, in which case the errors were subestimated. So we decided to maintain γ in the midterm. The main criticism is that, nevertheless, it was still chosen by hand.

As we have seen, we are able to ignore the factor γ and forget about any discretization error, but we have to find other means to account for the inheritance process of grid division. At the end of section 4.2.2 we saw that one possible way is to pass only a fraction of the parameters from one scale to the next. That is the same as first increasing the variance of the obtained posterior at some scale and then passing it as the prior of the next scale. We noticed that in the first few iterations, usually the first 2, we had to increase

the variances by large amounts. It would get smaller with larger grids, though. Although we have proposed one method in section 3.3.1, it did not predict large values for these parameter renormalization factors and the exact values are still unconstrained in theory. In any case, we did observe that under controlled assumptions the method works (in both 3 and 2 dimensions it behaves in the same general way) and we were able to predict the bulk content of some input charge distribution from a finite set of experimental data. It remains to analyse if the amount of detectors we have used is really necessary. The main question to answer concerning this issue and geo-neutrinos is: how many detectors are necessary to reconstruct Earth's inner distribution of radioactive elements at some predefined resolution? It is clear that the very own nature of inverse problems prevent us from drawing spectacular new results at scales of resolution larger than the number of input data; but there is always some information to gather, even in the presence of experimental errors. To approach this with the multigrid method, the first step is to analyse its behavior under a variable number of detectors.

A | Geo-Neutrino Oscillation and Spectra

A.1 Brief Review of Neutrino Oscillations in Vacuum

In quantum mechanics, oscillatory phenomena occur when a superposition of states is evolved in time by Schrödinger's equation. A simple example is the precession of the electron's magnetic moment in the presence of an external magnetic field. The same effect occurs if we assume that the neutrino flavor field $\nu_{\alpha L}$ is a superposition of neutrino mass fields ν_{iL} . We call the states $|\nu_{\alpha}\rangle$ produced by $\nu_{\alpha L}$ flavor eigenstates and states $|\nu_i\rangle$ produced by ν_{iL} mass eigenstates. The later are eigenstates of the free massive field ν_{iL} Hamiltonian

$$\mathcal{H} |\nu_i\rangle = E_i |\nu_i\rangle , \quad (\text{A.1})$$

where $E_i = \sqrt{|\vec{p}_i|^2 + m_i^2}$ is the energy eigenvalue, $\vec{p}_i$ the momenta and m_i the mass of state i . These fields are related by a unitary matrix U a way so that the flavor and mass eigenstates become related by

$$|\nu_{\alpha}\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle . \quad (\text{A.2})$$

To describe oscillations we make a number of assumptions that oversimplify the problem, allowing its treatment by basic quantum mechanics. For instance, we assume that:

- Neutrinos are produced and detected in the flavor basis;
- A neutrino's wave function is represented by a plane wave. This is unrealistic as we know particles are described by wave packets. A more complete deduction must treat neutrinos as wave packets;
- Mass eigenstates' energies are indistinguishable. Therefore, $E_i = E$, which implies that flavor eigenstates have definite energy E ;
- Propagation of all the mass eigenstates in space is given in the direction of a unique vector $\hat{\xi}$, that is $\vec{p}_i = p_i \hat{\xi}$. Therefore, if $\vec{x}$ is the propagation distance, $\vec{p}_k \cdot \vec{x} = p_i |\vec{x}|$.

Suppose a neutrino flavor eigenstate $|\nu_{\alpha}(\vec{x} = 0, t = 0)\rangle = |\nu_{\alpha}\rangle$ is produced in a charged weak process involving some charged lepton ℓ_{α} . We first evolve $|\nu_{\alpha}\rangle$ in space and time

with shift operators and then expand it in terms of the mass eigenstates as in (A.2):

$$|\nu_\alpha(\vec{x}, t)\rangle = e^{-iHt+i\vec{P}\cdot\vec{x}} |\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle e^{-iEt+i\vec{p}_i\cdot\vec{x}}. \quad (\text{A.3})$$

Now, since we are assuming that masses are small and energies are equal, we may approximate $p_k = \sqrt{E_k^2 - m_k^2} \approx E - m_k^2/2E$ so that the phase in (A.3) becomes $-i(Et - \vec{p}_k \cdot \vec{x}) \approx -iE(t - |\vec{x}|) - i|\vec{x}| m_k^2/2E$. If $|\nu_\alpha(\vec{x}, t)\rangle$ is then detected at coordinates $(\vec{x}, t)$ in a flavor state β described by $|\nu_\beta\rangle$, the probability of this experimental outcome is given by

$$|\langle\nu_\beta|\nu_\alpha(\vec{x}, t)\rangle|^2 = \left| \sum_{i,j} U_{\alpha i}^* U_{\beta j} \langle\nu_j|\nu_i\rangle e^{-iE(t-|\vec{x}|)-i|\vec{x}| m_k^2/2E} \right|^2. \quad (\text{A.4})$$

Assuming the orthonormalization of the mass eigenstates, $\langle\nu_j|\nu_i\rangle = \delta_{ij}$, and calling $L = |\vec{x}|$ we obtain the neutrino oscillation probability ¹,

$$\mathcal{P}_{\nu_\alpha \rightarrow \nu_\beta}(L, E) = \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\beta j}^* U_{\alpha j} \exp(-iL\Delta m_{ij}^2/2E), \quad (\text{A.5})$$

where $\Delta m_{ij}^2 = m_i^2 - m_j^2$ are the mass squared differences. Although there are three possible mass squared difference quantities $(\Delta m_{21}^2, \Delta m_{32}^2, \Delta m_{31}^2)$, only two are independent since $\Delta m_{21}^2 + \Delta m_{32}^2 - \Delta m_{31}^2 = 0$. In (A.5), the oscillatory behaviour is completely determined by the oscillation phase $\Phi_{kj} = -L\Delta m_{ij}^2/2E$. The oscillation amplitude is solely specified by the mixing matrix U . Since U is unitary, $\mathcal{P}_{\nu_\alpha \rightarrow \nu_\beta}$ is also unitary in both α and β .

For anti-neutrinos $\bar{\nu}_{\alpha L}$, described by the flavor eigenstates $|\bar{\nu}_\alpha\rangle$ and mass eigenstates $|\bar{\nu}_i\rangle$, the mixing is carried by the complex conjugate of U^* ,

$$|\bar{\nu}_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i} |\bar{\nu}_i\rangle. \quad (\text{A.6})$$

As a consequence, the oscillation probability of $\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta$ is similar to (A.5), the only difference being the complex conjugate sign on U ²,

$$\mathcal{P}_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta}(L, E) = \sum_{i,j} U_{\alpha i} U_{\beta i}^* U_{\beta j} U_{\alpha j}^* \exp(-iL\Delta m_{ij}^2/2E). \quad (\text{A.7})$$

In particular, the oscillation lengths are equal in both cases.

In order to further discuss oscillations, some words need to be said regarding the

¹We could also have arrived at (A.5) by assuming equal momenta for the mass eigenstates. However, in that case, the oscillation probability would be dependent on t instead of L and we would have to make the additional assumption that $t \approx L$, which is very reasonable as neutrinos are nearly massless.

²The observed symmetry in the flavor in (A.7) and (A.5) is not a mere coincidence; the identity $\mathcal{P}_{\nu_\alpha \rightarrow \nu_\beta} = \mathcal{P}_{\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha}$ is always valid as long as CPT symmetry is exact. CP symmetry interchanges neutrinos with anti-neutrinos T symmetry interchanges the initial and final states. CPT symmetry combines both effects of CP and T resulting in the exchange of $\nu_\alpha \rightarrow \nu_\beta$ to $\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha$. In particular, the survival probabilities are always equal, $\mathcal{P}_{\nu_\alpha \rightarrow \nu_\alpha} = \mathcal{P}_{\bar{\nu}_\alpha \rightarrow \bar{\nu}_\alpha}$.

parametrization of U , the unitary mixing matrix. It is usually referred as the PMNS matrix,

$$U = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}. \quad (\text{A.8})$$

The mass eigenstates' indexes $i = 1, 2, 3$ are labeled so to imply the decrescent ordering of the amount of ν_e in ν_i , $|\langle \nu_e | \nu_1 \rangle| > |\langle \nu_e | \nu_2 \rangle| > |\langle \nu_e | \nu_3 \rangle|$. Consequently, the PMNS matrix elements satisfy $|U_{e1}|^2 > |U_{e2}|^2 > |U_{e3}|^2$.

In general, a complex $N \times N$ matrix has $2N^2$ real independent parameters, of which N^2 are phases and N^2 are magnitudes of complex numbers. The unitary condition $UU^\dagger = \mathbf{1}$ provides N^2 independent equations of constraint and leaves room for only N^2 independent parameters. In two dimensions, it is very simple to work out the general formula for U since we have only 4 equations of constraint. After some appropriate reparametrization of the fields appearing in the Lagrangian, we find that U can be written as a simple $SO(2)$ rotation. In the three generations case the situation is more complicated. We have N^2 constraint equations which eliminate 9 parameters and the re-phasing of the fields allows the further elimination of 5 parameters. In this case, however, the rephasing leaves a common global phase δ . The standard parametrization of the 3×3 mixing matrix thus depends on 3 angles $\theta_{12}, \theta_{13}, \theta_{23}$ and one phase δ and is written as

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \quad (\text{A.9})$$

This choice of parametrization is useful because it assumes the simplest form for the so far measured quantities $|U_{e2}|^2$, $|U_{e3}|^2$ and $|U_{\mu 2}|^2$,

$$|U_{e3}|^2 = \sin^2 \theta_{13} \quad (\text{A.10})$$

$$|U_{\mu 2}|^2 = \sin^2 \theta_{23}(1 - |U_{e3}|^2) \approx \sin^2 \theta_{23} \quad (\text{A.11})$$

$$|U_{e2}|^2 = \sin^2 \theta_{12}(1 - |U_{e3}|^2) \approx \sin^2 \theta_{12}, \quad (\text{A.12})$$

where the last two approximations are due to the small value of $|U_{e3}|^2$.

In practice, the CP violation phase δ is invisible to most oscillation experiments and (A.9) can be approximated as the product of 3 $SO(3)$ rotations,

$$U = R^{23}R^{13}R^{12} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} \\ 0 & 1 & 0 \\ -s_{13} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (\text{A.13})$$

Neutrino oscillations are known today as a fact. Evidence comes from a great variety of experimental setups, ranging from energy scales of MeV's to many GeV's and baseline distances of km's to a.u.'s. Apart from a few recent anomalies, the standard theory of oscillations accounts, in the simplest and most precise way, for all the observed data.

Our knowledge of neutrino oscillations has been constructed over the years based on two so far known scales of L/E , corresponding to two distinct values of Δm_{ij}^2 : Δm_{21}^2 and Δm_{31}^2 . Δm_{21}^2 , also called Δm_{SOL}^2 , accounts for values of L/E of neutrinos coming

from the sun. $|\Delta m_{31}|^2$, or Δm_{ATM}^2 , corresponds to values of L/E of atmospheric neutrinos, secondary products of cosmic rays that hit Earth's atmosphere. Since $\Delta m_{\text{SOL}}^2 \ll \Delta m_{\text{ATM}}^2$, two possible mass orderings emerge: the normal hierarchy, where $m_3 \gg m_2 > m_1$, and the inverted hierarchy, where $m_3 \ll m_1 < m_2$.

In the 3 flavor scenario, the oscillation probability of these two scales decouple due to the small value of $|U_{e3}|$, behaving effectively as 2 flavor oscillations. Thus, the effective 2 flavor transition probabilities for the solar and atmospheric cases can be approximated by taking $\theta_{21} = \theta_{\text{SOL}}$ and $\theta_{31} = \theta_{\text{ATM}}$. So, in the absence of information of one of the decoupled system, one can still make predictions about the other in a certain degree of accuracy. Recent global analysis of the neutrino oscillation parameters [Fogli et al., 2012] give the following 1σ estimates for the normal hierarchy:

$$\Delta m_{21}^2 = 7.54_{-0.22}^{+0.26} \times 10^{-5} \text{eV}^2, \quad (\text{A.14})$$

$$\Delta m_{31}^2 = 2.43_{-0.10}^{+0.06} \times 10^{-3} \text{eV}^2, \quad (\text{A.15})$$

$$\sin^2 \theta_{12} = 0.307_{-0.016}^{+0.018}, \quad \sin^2 \theta_{23} = 0.386_{-0.021}^{+0.024}, \quad \sin^2 \theta_{13} = 0.024_{-0.25}^{+0.025} \quad (\text{A.16})$$

A.2 Geo-Neutrino Oscillation

A.2.1 Interactions with Matter

While propagating in matter, neutrinos are subject to interactions that continuously modify the mixing parameters. These are essentially $W^\pm$ and Z interactions by scattering with the matter constituents (protons, electrons and neutrons). The total effective Hamiltonian that describes these scattering processes at low energies is given by the sum of a charged current part $\mathcal{H}_{\text{eff}}^{\text{cc}}$ and a neutral current part $\mathcal{H}_{\text{eff}}^{\text{nc}}$:

$$\mathcal{H}_{\text{eff}}^{\text{cc}} = \frac{G_F}{\sqrt{2}} \bar{\nu}_e(x) \gamma^\mu (1 - \gamma^5) e(x) \bar{e}(x) \gamma_\mu (1 - \gamma^5) \nu_e(x), \quad (\text{A.17})$$

$$\mathcal{H}_{\text{eff}}^{\text{nc}} = \frac{G_F}{\sqrt{2}} \sum_{\alpha=e,\mu,\tau} \bar{\nu}_\alpha(x) \gamma^\mu (1 - \gamma^5) \nu_\alpha(x) \sum_{f=e,p,n} \bar{f}(x) \gamma_\mu (g_V^f - g_A^f \gamma^5) f(x), \quad (\text{A.18})$$

where g_V^f and g_A^f are the vector and axial coupling constants related to each spin 1/2 matter particle f . (A.17) can be further simplified by exchanging the fields ν and e through a Fierz transformation,

$$\mathcal{H}_{\text{eff}}^{\text{cc}} = \frac{G_F}{\sqrt{2}} \bar{\nu}_e(x) \gamma^\mu (1 - \gamma^5) \nu_e(x) \bar{e}(x) \gamma_\mu (1 - \gamma^5) e(x). \quad (\text{A.19})$$

Electronic neutrinos are able to couple and interact with electrons by charged weak currents, as in (A.17). But electrons do not couple to muon or tau neutrinos by neither W^+ nor W^- ; they do couple, however, by neutral currents, as in (A.18).

In order to describe how these interactions in matter affect the propagation of neutrinos, we derive from the total effective low-energy weak Hamiltonian $\mathcal{H}_{\text{eff}} = \mathcal{H}_{\text{eff}}^{\text{cc}} + \mathcal{H}_{\text{eff}}^{\text{nc}}$ an effective potential V_{eff} that accounts for all the possible values of spin h_f and momentum q_f that the matter constituents f can have at the medium's temperature T and we subject

neutrino states $|\bar{\nu}_e\rangle$ to the effects of V_{eff} ,

$$V_{\text{eff}} = \sum_{f=e,n,p} \int d^3x \int d^3q_f f(E_f, T) \frac{1}{2} \sum_{h_f=\pm 1} \langle p, e, n, \bar{\nu}_e | \mathcal{H}_{\text{eff}} | p, e, n, \bar{\nu}_e \rangle , \quad (\text{A.20})$$

where $f(E_f, T)$ is the Fermi-Dirac distribution at a temperature T and E_f is the energy of the spin 1/2 particle f . If we consider that the medium is homogeneous, A.20 gives

$$V_{\text{eff}}^{cc} \approx \begin{cases} G_F n_e m_{\nu_\alpha}^2 / 4E_\nu^2 , & h_\nu = +1 \\ \sqrt{2} G_F n_e , & h_\nu = -1 \end{cases} , \quad (\text{A.21})$$

where the quantity $\frac{1}{V} \int d^3q_e f(E_e, T) = n_e$ is the electron number density of the medium. Analogously, for the neutral current part of V_{eff} we obtain for any neutrino flavor α ,

$$V_{\text{eff},f}^{nc} \approx \begin{cases} G_F n_f g_V^f m_{\nu_\alpha}^2 / 4E_\nu^2 , & h_\nu = +1 \\ \sqrt{2} G_F n_f g_V^f , & h_\nu = -1 \end{cases} . \quad (\text{A.22})$$

That is, for ultra-relativistic left-handed neutrinos of flavor α , for which its helicity h_ν coincides with its chirality -1 , the effective potential in matter is given by

$$V_{\text{eff},\alpha} \approx \sqrt{2} G_F (\delta_{\alpha e} n_e + g_V^e n_e + g_V^p n_p + g_V^n n_n) = \sqrt{2} G_F \left(\delta_{\alpha e} n_e - \frac{1}{2} n_n \right) , \quad (\text{A.23})$$

where n_p and n_n are the proton and neutron number density in matter. We see that the electronic neutral current part contribution cancels the proton contribution due to their opposite charge.

For right-handed ultra-relativistic anti-neutrinos ($h_\nu = +1$), the potential acquires a negative sign due to their opposed helicity:

$$\bar{V}_{\text{eff},\alpha} \approx -\sqrt{2} G_F \left(\delta_{\alpha e} n_e - \frac{1}{2} n_n \right) . \quad (\text{A.24})$$

A.2.2 Oscillations in Matter

Suppose now a neutrino state is created at $t = 0$ and $L = 0$ in the initial state $|\nu_\alpha\rangle$. The evolution of the flavor transition amplitude $\mathcal{A}_{\alpha \rightarrow \beta}(t) = \langle \nu_\beta | \nu_\alpha(t) \rangle$ in the presence of a medium's potential V_α is given by Schrödinger's equation with hamiltonian $H = H_0 + V_\alpha$, where H_0 is the free particle's hamiltonian,

$$i \frac{d}{dt} \langle \nu_\beta | \nu_\alpha(t) \rangle = \langle \nu_\beta | H | \nu_\alpha \rangle = \sum_\sigma \left(\sum_k U_{\beta k} U_{\sigma k} E_k + \delta_{\beta \sigma} V_\sigma \right) \langle \nu_\sigma | \nu_\alpha \rangle . \quad (\text{A.25})$$

To derive a simpler evolution equation, we assume neutrinos to be ultra-relativistic $x \approx t$. We further assume that all the three mass eigenstates have equal momenta, so that $p_k = p \approx E$ for all k , where E is the energy neglecting the mass term. Therefore,

$E_k \approx E + m_k^2/2E$. Denoting $\mathcal{A}_{\alpha\beta}(x) = \langle \nu_\beta | \nu_\alpha(x) \rangle$, equation (A.25) becomes

$$i \frac{d}{dx} \mathcal{A}_{\alpha\beta}(x) = \sum_{\sigma} \left(\sum_k U_{\beta k} \frac{\Delta m_{k1}^2}{2E} U_{\sigma k}^* + \delta_{\beta\sigma} \delta_{\sigma e} V_{cc} \right) \mathcal{A}_{\alpha\sigma}(x), \quad (\text{A.26})$$

where we have dropped the term $(E + m_1^2/2E + V_{NC}) \mathcal{A}_{\alpha\beta}(x)$ because it can be eliminated by rephasing $\mathcal{A}_{\alpha\beta}$ properly. In matrix form, (A.26) becomes

$$i \frac{d}{dx} \begin{pmatrix} \mathcal{A}_{\alpha e} \\ \mathcal{A}_{\alpha \mu} \\ \mathcal{A}_{\alpha \tau} \end{pmatrix} = \frac{1}{2E} \left[U \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Delta m_{21}^2 & 0 \\ 0 & 0 & \Delta m_{31}^2 \end{pmatrix} U^T + A_{cc} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] \begin{pmatrix} \mathcal{A}_{\alpha e} \\ \mathcal{A}_{\alpha \mu} \\ \mathcal{A}_{\alpha \tau} \end{pmatrix}, \quad (\text{A.27})$$

where $A_{cc} = 2EV_{cc}$. We refer to the amplitude vectors in (A.27) as $\mathcal{A}_\alpha^F$, to suggest its flavor basis content.

In the case of anti-neutrinos, because the mixing relation (A.6) involves U instead of U^* , the first matrix U in equation (A.27) is replaced by U^* and the second one by U^T . Taking in account the choice of parametrization (A.13), the evolution equation for anti-neutrinos coincides with (A.27).

5.2.3 Oscillations inside the Earth

Due to interactions inside the Earth, the measured flux of incoming $\bar{\nu}_e$ should in principle deviate from the expected vacuum value. However, the exact equations of neutrino oscillation in matter are complicated. To simplify the description, we analyse matter effects in geo-neutrino propagation assuming PREM's density profile [Dziewonski and Anderson, 1981]. In PREM, the electronic density number of the Earth n_e^{Earth} (figure 5.1a) becomes a single variable function of the distance from the center and the matter potential (figure 5.1b) becomes simply

$$A_{cc}^{\text{Earth}} \approx 2E_{\bar{\nu}} \times 7.63 \times 10^{-14} n_e^{\text{Earth}} \frac{\text{eVcm}^3}{N_A}. \quad (\text{5.28})$$

Due to the small value of n_e^{Earth} throughout the Earth, A_{cc} is comparatively small to Δm_{12}^2 and Δm_{31}^2 in the energy range of geo-neutrinos.

(a) Earth's electronic number density given by PREM.

(b) Effective potential for 2 MeV geo-neutrinos assuming PREM's electronic density profile.

Using the parametrization of the mixing matrix described in (A.13), we may redefine the flavor transition amplitude $\mathcal{A}_\alpha^F$ of (A.27) to $\widetilde{\mathcal{A}}_\alpha^F = R^{13T} R^{23T} \mathcal{A}_\alpha^F$ and derive a simpler form of the evolution equation in matter for electronic anti-neutrinos,

$$i \frac{d}{dx} \widetilde{\mathcal{A}}_e^F = \frac{1}{2E} \begin{pmatrix} s_{12}^2 \Delta m_{12}^2 + c_{31}^2 A_{cc} & c_{12} s_{12} \Delta m_{12}^2 & -c_{13} s_{13} A_{cc} \\ c_{12} s_{12} \Delta m_{12}^2 & c_{12}^2 \Delta m_{12}^2 & 0 \\ -c_{13} s_{13} A_{cc} & 0 & \Delta m_{31}^2 + s_{13}^2 A_{cc} \end{pmatrix} \widetilde{\mathcal{A}}_e^F. \quad (5.29)$$

where we have re-phased the amplitude vector $\widetilde{\mathcal{A}}_e^F$ by a factor $\exp -i (\Delta m_{12}^2 + \cos \theta_{13} + A_{cc})$ to simplify the equation. Because A_{cc} is small and varies very little, equation (5.29) implies that the third eigenvalue of the Hamiltonian is simply Δm_{31}^2 and the third eigenstate decouples from the evolution equation. We are left with a two flavor approximation with eigenvalues $\pm \Delta m_M^2$ and mixing angle θ_M , shown in figures 5.2a and 5.2b, given by

$$\Delta m_M^2 = \sqrt{(\Delta m^2 \cos 2\theta - \cos^2 \theta_{13} A_{cc})^2 + (\Delta m^2 \sin 2\theta)^2}, \quad (5.30)$$

$$\sin 2\theta_M = \frac{\sin 2\theta_{12} \Delta m_{12}^2}{\Delta m_M^2}, \quad \cos 2\theta_M = \frac{\cos 2\theta_{13} \Delta m_{12}^2 - \cos^2 \theta_{13} A_{cc}}{\Delta m_M^2}. \quad (5.31)$$

(a) Effective mass squared differences in Earth's interior for the 1-2 sector. The horizontal line represents Δm_{12}^2

(b) Effective mixing angle in Earth's interior.

The adiabaticity parameter in this case is $\gamma \sim 10^5 \gg 1$, except possibly at the density discontinuity zones, where θ_M also has discontinuities. At each density slab, however, the oscillation can be considered to be adiabatic. We will assume that these density transition zones have some considerable extent, so that, in practice, evolution is always adiabatic. In this simplified scenario, all the three flavor-mass amplitudes $\mathcal{A}_e^F = U_M \mathcal{A}_e^M$ have their own evolution equation. So if we assume the initial state to be an electronic anti-neutrino $(\mathcal{A}_e^F)_\alpha = \delta_{e\alpha}$,

$$\mathcal{A}_e^M(0) = U_M^T R^{13T} R^{23T} \mathcal{A}_e^F(0) = \begin{pmatrix} \cos \theta_M^i c_{13} \\ -\sin \theta_M^i c_{13} \\ s_{13} \end{pmatrix}, \quad (5.32)$$

then the exact survival probability becomes

$$\begin{aligned} \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} &= |\mathcal{A}_{ee}^F|^2 = |(R^{23} R^{13} U_M \mathcal{A}_e^F)_e|^2 \\ &= |c_{13}^2 \cos \theta_M^f \cos \theta_M^i e^{-i\Delta_M} + c_{13}^2 \sin \theta_M^f \sin \theta_M^i e^{i\Delta_M} + s_{13}^2 e^{-i\Delta_{31}}|^2, \end{aligned} \quad (5.33)$$

where $\Delta_{31} = \Delta m_{31}^2 x / 4E$, $\Delta_M = \int_0^r dr' \Delta m_M^2(r') / 4E$, θ_M^i is the matter mixing angle at the production point and θ_M^f at the detection point.

5.2.4 Approximations

Due to the production and detection processes, detectable geo-neutrinos are restricted to a small range of energy $E \sim 2$ MeV. The associated oscillation lengths for this value of energy are $L_{12}^{osc} = 66$ km and $L_{31}^{osc} = 2$ km, from which we see that oscillations controlled by Δm_{31}^2 require a great contribution from geo-neutrinos produced very near the detector to be visible. If we assume that this is not the case, 3-1 oscillations are washed out by the detector's resolution. Also, we may neglect $\sin^4 \theta_{13} \sim 10^{-4}$ and approximate $\cos^4 \theta_{13} \sim (1 - 10^{-2})^2 \approx 1$, so that

$$\mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} = \frac{1}{2} + \frac{1}{2} \cos 2\theta_M^f \cos 2\theta_M^i + \frac{1}{2} \sin 2\theta_M^f \sin 2\theta_M^i \cos \Delta_{12}. \quad (5.34)$$

Similarly, we may also take $\cos 2\theta_M^i \cos 2\theta_M^f \approx \cos^2 2\theta_{12}$ and $\sin 2\theta_M^i \sin 2\theta_M^f \approx \sin^2 2\theta_{12}$, which will only contribute with small deviations to the amplitude of the oscillation probability,

$$\mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} = 1 - \frac{1}{2} \sin^2 2\theta_{12} + \frac{1}{2} \sin^2 2\theta_{12} \cos \Delta_{12}. \quad (5.35)$$

Figures 5.3a and 5.3b show the effects of the last approximations, comparing the approximated survival probabilities with the exact form (5.33).

(a) The exact survival probability (5.33) (curve in red) in comparison to the approximated oscillation probability (5.34) (in black) for 2 MeV's neutrinos created at 4000 km from the center of the Earth and detected at the surface. If the wiggly behavior of the exact probability cannot be captured by the experiment's precision, it may be approximated by the curve in black.

(b) Comparison between probabilities (5.34) and (5.35), where matter angles have been approximated by the vacuum ones, for 2 MeV's neutrinos created at 4000 km from the center of the Earth and detected at the surface.

Furthermore, we may simplify (5.35) even more by taking its average over the energy spectrum,

$$\langle \mathcal{P}_{\bar{\nu}_e \rightarrow \bar{\nu}_e} \rangle = 1 - \frac{1}{2} \sin^2 2\theta_{12} = 0.58 . \quad (5.36)$$

The effects of this last approximation in the geo-neutrino flux was studied in [Sanshiro, 2005]. For distances greater than ~ 100 km, the averaging amounts to an almost perfect approximation of the flux.

5.3 Spectra

The anti-neutrino flux depends explicitly on the energy spectrum of the produced $\bar{\nu}_e$ in the β^- process that generated it. In order to calculate the decay spectrum $d\Gamma/dE_e = f_e(E_e)$ of a particular nucleus of atomic number Z , we assume that it follows the universal shape of (allowed) beta decay in terms of the emitted electron's energy E_e , mass m_e and momentum $p_e = \sqrt{E_e^2 - m_e^2}$,

$$f_e(E_e) = \frac{1}{N} F(Z, E_e) (E_e^{\text{MAX}} - E_e)^2 p_e E_e , \quad (5.37)$$

where $E_e^{\text{MAX}} = Q + m_e$ is the maximal energy electrons can acquire (Q being the maximum kinetic energy) and $F(Z, E_e)$ is the Fermi correction that accounts for Coulomb interactions between the final nucleus and the emitted electron,

$$F(Z, E_e) = p_e^{2\gamma-2} e^{\pi y(E_e)} |\Gamma(\gamma + iy(E_e))|^2 , \quad (5.38)$$

In (5.38), α is the fine-structure constant and $y(E_e) = \alpha Z E_e / p_e$, $\gamma = \sqrt{1 - (\alpha Z)^2}$. N is a normalization constant given by

$$N = \int_{m_e}^{E_e^{\text{MAX}}} f_e(E_e) dE_e . \quad (5.39)$$

In order to obtain the spectrum in terms of the anti-neutrino energy E_ν , we assume energy conservation and neglect both the proton's recoil and the neutrino mass. In the limit case, when the electron is emitted with zero kinetic energy, the anti-neutrino has its maximum possible energy $E_\nu^{\text{MAX}} \approx Q$, and we may write $E_\nu + E_e = m_e + E_\nu^{\text{MAX}}$. On the other hand, the neutrino may also be emitted with zero kinetic energy. If we neglect its mass, $E_\nu^{\text{MAX}} + m_e = E_e^{\text{MAX}}$ and we may write the $\bar{\nu}_e$ energy spectrum by reflecting the electron's energy by $E_\nu = E_e^{\text{MAX}} - E_e$,

$$f_\nu(E_\nu) = f_e(E_e)|_{E_e=E_e^{\text{MAX}}-E_\nu} \quad (5.40)$$

$$= \frac{1}{N} E_\nu^2 (E_e^{\text{MAX}} - E_\nu) ((E_e^{\text{MAX}} - E_\nu)^2 - m_e^2)^{\gamma-1/2} e^{\pi y} |\Gamma(\gamma + iy)|^2 . \quad (5.41)$$

It should be noted that most of the geo-neutrino producing elements, shown in table 5.1, actually decay by forbidden β^- transitions. Therefore, expression (5.41) will serve us only as an rough approximation to the energy spectra of geo-neutrinos.

^{238}U transitions	$P_{i \rightarrow j}$	$E_{max}^{\bar{\nu}_e}$ (KeV)	^{232}Th transitions	$P_{i \rightarrow j}$	$E_{max}^{\bar{\nu}_e}$
$^{234}\text{Pa}^m \rightarrow ^{234}\text{U}$	0,9984	2268,92	$^{228}\text{Ac} \rightarrow ^{228}\text{Th}$	1	2069,24
$^{214}\text{Bi} \rightarrow ^{214}\text{Po}$	0,9998	3272	$^{212}\text{Bi} \rightarrow ^{212}\text{Po}$	0,6406	2254

Table 5.1: β transitions in the ^{238}U and ^{232}Th decay chain responsible for the currently observed geo-neutrinos (elements with non-negligible occurrence probability and with $E_{max}^{\bar{\nu}_e}$ greater than 1.806 MeV). $P_{i \rightarrow j}$ gives the branching ratio and $E_{max}^{\bar{\nu}_e}$ gives the maximum energy of the emitted anti-neutrino.

The total spectrum is obtained by summing up the spectra of the individual transitions,

$$f(E) = \sum_{ij} P_{i \rightarrow j} \sum_k I_{i \rightarrow j; k} f_{i \rightarrow j; k}(E) . \quad (5.42)$$

The index $i \rightarrow j$ indicates the process of transformation of a nucleus i into a j . An extra index k may be added if the process $i \rightarrow j$ can occur in more than one way, possibly involving different energy transitions. $I_{i \rightarrow j; k}$ denotes the relative intensity of the different transitions $i \rightarrow j; k$ [Firestone and Shirley, 1999] and $P_{i \rightarrow j}$ is the probability of the process $i \rightarrow j$ to occur.

Figure 5.4: ^{232}Th and ^{238}U chain normalized energy spectra of the transitions listed in table 5.1.

Bibliography

- [Abe et al., 2008] Abe, S. et al. (2008). Precision Measurement of Neutrino Oscillation Parameters with KamLAND. *Physical Review Letters*, 100(22):221803.
- [Araki et al., 2005] Araki, T. et al. (2005). Experimental investigation of geologically produced antineutrinos with kamLAND. *Nature*, 436:499–503.
- [Autiero et al., 2007] Autiero, D. et al. (2007). Large underground, liquid based detectors for astro-particle physics in europe: scientific case and prospects. <http://arxiv.org/abs/0705.0116>.
- [Barabanov et al., 2009] Barabanov, I., Novikova, G., Sinev, V., and Yanovich, E. (2009). Research of the natural neutrino fluxes by use of large volume scintillation detector at baksan. <http://arxiv.org/abs/0908.1466>.
- [Bellini et al., 2010] Bellini, G. et al. (2010). Observation of geo-neutrinos. *Phys. Lett. B*, 687(45).
- [Bellini et al., 2013] Bellini, G. et al. (2013). Measurement of geo-neutrinos from 1353 days of Borexino. *Physics Letters B*, 722(45):295 – 300.
- [Caticha, 2012] Caticha, A. (2012). Entropic inference and the foundations of physics. <http://www.albany.edu/physics/ACaticha-EIFP-book.pdf>. Book under preparation.
- [Caticha,] Caticha, N. Sistemas de processamento de informação. Book under preparation.
- [Cox, 2001] Cox, R. T. (2001). *Algebra of Probable Inference*. The Johns Hopkins University Press.
- [da Rocha Amaral et al., 2004] da Rocha Amaral, S., Rabbani, S. R., and Caticha, N. (2004). Multigrid priors for a bayesian approach to fMRI. *Neuroimage*, 23(2):654–662.
- [da Rocha Amaral et al., 2007] da Rocha Amaral, S., Rabbani, S. R., and Caticha, N. (2007). BOLD response analysis by iterated local multigrid priors. *Neuroimage*, 36(2):361–369.
- [da Silva Barbosa, 2011] da Silva Barbosa, L. (2011). Inclusão de MRI e Informação Multigrid a Priori para inferência Bayesiana de fontes de M/EEG. Master’s thesis, Instituto de Física, Universidade de São Paulo.

- [Davies and Davies, 2010] Davies, J. H. and Davies, D. R. (2010). Earth’s surface heat flux. *Solid Earth*, 1(1):5–24.
- [Dye, 2012] Dye, S. T. (2012). Geoneutrinos and the radioactive power of the Earth. *Reviews of Geophysics*, 50(3).
- [Dziewonski and Anderson, 1981] Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297 – 356.
- [Eder, 1966] Eder, G. (1966). Terrestrial neutrinos. *Nuclear Physics*, 78:657–662.
- [Fiorentini et al., 2007] Fiorentini, G., Lissia, M., and Mantovani, F. (2007). Geoneutrinos and Earth’s interior. *Phys. Rept.*, 453:117–172.
- [Firestone and Shirley, 1999] Firestone, R. B. and Shirley, V. S. (1999). *Table of Isotopes*. John Wiley & Sons, Inc., 8 edition.
- [Fogli et al., 2012] Fogli, G. L., Lisi, E., Marrone, A., Montanino, D., Palazzo, A., and Rotunno, A. M. (2012). Global analysis of neutrino masses, mixings, and phases: Entering the era of leptonic CP violation searches. *Phys. Rev. D*, 86(1):013012.
- [Gando et al., 2011] Gando, A. et al. (2011). Partial radiogenic heat model for earth revealed by geoneutrino measurements. *Nature Geoscience*, 4:647–651.
- [Gando et al., 2013] Gando, A. et al. (2013). Reactor on-off antineutrino measurement with KamLAND. *Phys. Rev. D*, 88:033001.
- [Giunti and Kim, 2007] Giunti, C. and Kim, C. W. (2007). *Fundamentals of Neutrino Physics and Astrophysics*. Oxford University Press.
- [Jaupart et al., 2007] Jaupart, C., Labrosse, S., and Mareschal, J.-C. (2007). 7.06 - Temperatures, Heat and Energy in the Mantle of the Earth. In Schubert, G., editor, *Treatise on Geophysics*. Elsevier.
- [Jaynes, 2003] Jaynes, E. T. (2003). *Probability Theory, The Logic of Science*. Cambridge University Press.
- [Krauss et al., 1984] Krauss, L. M., Glashow, S. L., and Schramm, D. N. (1984). Antineutrino astronomy and geophysics. *Nature*, 310:191–198.
- [Learned et al., 2007] Learned, J. G., Dye, S. T., and Pakvasa, S. (2007). Hanohano: A deep ocean anti-neutrino detector for unique neutrino physics and geophysics studies. In *Proceedings of the Twelfth International Workshop on Neutrino Telescopes*.
- [Loredo, 2003] Loredo, T. J. (2003). Bayesian Adaptive Exploration. In *Bayesian Inference And Maximum Entropy Methods In Science And Engineering: 23rd International Workshop*, volume 707 of *AIP Conference Proceedings*, pages 330–346. American Inst. of Physics.
- [Lozza, 2012] Lozza, V. (2012). Scintillator phase of the sno+ experiment. <http://arxiv.org/abs/1201.6599>.

- [Machado et al., 2012] Machado, P. A. N., Mühlbeier, T., Nunokawa, H., and Funchal, R. Z. (2012). Potential of a neutrino detector in the ANDES underground laboratory for geophysics and astrophysics of neutrinos. *Phys. Rev. D*, 86:125001.
- [Marx, 1969] Marx, G. (1969). Geophysics by neutrinos. *Czech. J. Phys. B*, 19:1471–1479.
- [McDonough, 2003] McDonough, W. (2003). 2.15 - compositional model for the earth’s core. In Holland, H. D. and Turekian, K. K., editors, *Treatise on Geochemistry*, pages 547 – 568. Pergamon.
- [Mikheyev and Smirnov, 1985] Mikheyev, S. P. and Smirnov, A. Y. (1985). Resonance enhancement of oscillations in matter and solar neutrino spectroscopy. *Sov. J. Nucl. Phys.*, 42:913917.
- [Miller and Karl,] Miller, E. L. and Karl, W. C. Fundamentals of Inverse Problems. Lecture notes from Northeastern University course on inverse problems.
- [Sanshiro, 2005] Sanshiro, E. (2005). *Neutrino Geophysics and Observation of Geo-Neutrinos at KamLAND*. PhD thesis, Tohoku University.
- [Sramek et al., 2013] Sramek, O., McDonough, W. F., Kite, E. S., Lekic, V., Dye, S. T., and Zhong, S. (2013). Geophysical and geochemical constraints on geoneutrino fluxes from Earth’s mantle. *Earth and Planetary Science Letters*, 361:356 – 366.
- [Tajima, 2013] Tajima, S. (2013). Defining statistical perceptions with an empirical bayesian approach. *Phys. Rev. E*, 87:042707.
- [Tan and Fox,] Tan, S. M. and Fox, C. Lecture Notes from Physics 707 - Inverse Problems. Course given at the University of Auckland.
- [Tolich et al., 2007] Tolich, N., Chan, Y.-D., Currat, C., Fujikawa, B., Henning, R., Lesko, K., Poon, A., Decowski, M., Wang, J., and Tolich, K. (2007). A geoneutrino experiment at homestake. In Dye, S. T. ., editor, *Neutrino Geophysics: Proceedings of Neutrino Sciences 2005*, pages 229–240. Springer New York.
- [Wilde et al., 2001] Wilde, S. A., Valley, J. W., Peck, W. H., and Graham, C. M. (2001). Evidence from detrital zircons for the existence of continental crust and oceans on the earth 4.4 gyr ago. *Nature*, 409:175–178.
- [Wolfenstein, 1978] Wolfenstein, L. (1978). Neutrino oscillation in matter. *Phys. Rev. D*, 17(9):2369–2374.
- [Wurm et al., 2010] Wurm, M. et al. (2010). The physics potential of the lena detector. <http://arxiv.org/abs/1004.3474>.