Republic Of Algeria Democratic And People's Ministry Of Higher Education And Scientific Reseach

University of KASDI Merbah - Ouargla Faculty of Mathematics and Material Science Department of Physics



Master Academic Thesis

Domain: Physics Specialty: Radiation physics

Presented by :

## BENHADJIRA Abderrahmane

Theme :

# Judd-Ofelt parameters : Bayesian inference and deep learning approach

# 17/06/2021 in front of the jury members

$\mathbf{Pr}$	DOUIS Said	President	University of Quargla
Dr	BENSAVAH Abdallah	Inry	University of Ouargla
$D_{n}$		Jury	University of Quargla
Dr. D	AIADI Uussaina	Jury C ·	University of Ouargia
Dr.	BENTOUILA Omar	Supervisor	University of Ouargla
Pr.	AIADI Kamel Eddin	Co-suppervisor	University of Ouargla

# **Acknowledgements**

Firstly praise to be to Allah for his blessings, i thank Allah that i was able to finish this work.

I would like to thank my supervisor Dr.BENTOUILA Omar and co-supervisor Pr.AIADI Kamel Eddin for their guidance, the considerable time spent on consultations, proof reading, and for inspiring me to start the work on this thesis in the first place.

I would like to acknowledge Pr.BENSAYAH Abdallah and BOURENANE Mohamed Adem, for implementation support, i really appreciate their help.

B.Abderrahmane.

# Dedication

Firstly, Praise be to Allah for his blessings, I dedicate this work to the most valuable persons in my life, my mother and my father, Thank you for all the support you have given to me since i was a child. to my dear sisters and brothers, to all of my family, to my close friends B.Abderrahmane.

# Abstract

Dans ce travail, deux nouvelles approches pour déterminer les paramètres de Judd-Ofelt pour les ions de terres rares dans les verres ont été présentées : l'inférence bayésienne et l'apprentissage profond. Ces approches sont un cadre alternatif de l'ajustement conventionnel des moindres carrés des données expérimentales utilisées actuellement. La technique d'inférence bayésienne a été testée avec succès, et les paramètres de Judd-Ofelt estimés des verres halogéno-phosphates dopés à l'holmium et à l'erbium montrent un bon accord avec la méthode des moindres carrés. La deuxième technique est un perceptron multicouche de régression non linéaire (MLP) qui a été utilisé pour ajuster une matrice de composition de verres de tellurite dopés à l'erbium avec leurs paramètres Judd-Ofelt correspondants. Nous réussissons à détecter la corrélation avec un intervalle d'erreur raisonnable, en général la méthode a réussi et peut être encore améliorée pour prédire les paramètres de Judd-Ofelt.

**keywords:** Judd-Ofelt theory; Rare earth doped glasses; Bayesian Inference; Deep Learning; prediction.

# Résumé

Dans ce travail, deux nouvelles approches pour déterminer les paramètres de Judd-Ofelt pour les ions de terres rares dans les verres ont été présentées : l'inférence bayésienne et l'apprentissage profond, ces approches sont un cadre alternatif de l'ajustement conventionnel des moindres carrés des données expérimentales utilisées actuellement, La technique d'inférence bayésienne a été testée avec succès, et les paramètres de Judd-Ofelt estimés des verres halogéno-phosphates dopés avec l'Holmium et l'Erbium montrent un bon accord avec la méthode des moindres carrés. La deuxième technique est un perceptron multicouche de régression non linéaire (MLP) qui a été utilisé pour ajuster une matrice de composition de verres de tellurite dopés à l'erbium avec leurs paramètres Judd-Ofelt correspondants. Nous réussissons à détecter la corrélation avec un intervalle d'erreur raisonnable, en général, la méthode a réussi et peut être encore améliorée pour prédire les paramètres de Judd-Ofelt.

**Mots-clés:** Théorie de Judd-Ofelt ; Verres dopés terres rares; Inférence bayésienne ; Apprentissage profond; Prédiction

ملخص

في هذا العمل، تم تقديم طريقتين جديدتين لتحديد معاملات جودأوفلت لأيونات العناصر الترابية النادرة في الزجاج: الاستدلال البايزي والتعلم العميق. هاتان الطريقتان يمكن اعتبارهما كبدائل لطريقة المربعات الصغرى المعتمدة على البيانات التجريبية المستخدمة حاليًا. تم اختبار تقنية الاستدلال البايزي بنجاح، وتُبين معاملات جودأوفلت المتحصّل عليها لزجاج هاولوجينوفوسفاتي مُطعّم بلهوليوم والإربيوم توافقا جيدا مع طريقة المربعات الصغرى. الطريقة الثانية هي الاتحدار غير الخطي متعدد الطبقات التي تم استخدامها لتوفيق مصفوفة من مكونات زجاج تيلوريت مطعم بالإربيوم مع معاملات جود أوفلت المقابلة لها. نجحنا في اكتشاف الارتباط ضمن مجال معقول للخطأ، بشكل عام كانت الطريقة ناجة و يكمن ان تحسن اكثر من اجل التنبؤ بمعاملات جودأوفلت.

**الكلمات الفتاحية :** نظرية جود أوفلت؛ الزجاج المُطعّم بالعناصر الترابية النادرة؛ الاستدلال البايزي؛ التعلم العميق؛ التنبؤ

# Contents

Acknow	wledge	ments	<b>2</b>		
Dedica	tion		4		
Abstra	$\mathbf{ct}$		<b>5</b>		
Genera	al Intro	oduction	14		
1	Introd	uction	14		
2	Thesis	scope	14		
3	Thesis	structure	15		
Chapte	er 1:	Rare Earths And Judd Ofelt Theory	16		
1	Introd	uction	16		
2	Rare I	Earths	16		
	2.1	Rare Earths ions	17		
	2.2	Rare Earths Spectroscopy	18		
		2.2.1 Energy Levels	18		
		2.2.2 Schrodinger equation solution in central field approximation	19		
		2.2.3 LS and Intermediate Coupling	21		
		2.2.4 Crystal field approximation	24		
3	Judd	Ofelt theory	26		
	3.1	Electric dipole approximation	28		
	3.2	Magnetic dipole approximation	29		
	3.3	Judd-Ofelt parameters calculation	29		
	3.4	Conclusion	30		
Chapte	er 2:	Bayesian Inference	31		
1	Introd	uction	31		
2	Conditional Probability				
3	Bayes	theorem	32		
4	Linear regression				

	4.1	Conventional linear regression	34
	4.2	Bayesian linear regression	36
		4.2.1 Prior	37
		4.2.2 Posterior	38
		4.2.3 Markov Chain Monte Carlo simulation	39
5	Concl	usion	40
Chapt	er 3:	Judd Ofelt parameters prediction	41
1	Introd	luction	41
	1.1	Regression	41
	1.2	States of the art	42
2	Neura	I networks and deep learning	42
	2.1	Perceptron(artificial neuron)	43
	2.2	Multilayer perceptron	44
3	Traini	ing Neural neural networks	45
	3.1	Loss function	47
	3.2	Gradient decent	47
	3.3	Backpropagation	48
	3.4	Regularization	50
		3.4.1 Dropout	50
4	conclu	usion	51
Chapt	er 4:	Results and discussion	52
1	Introd	luction	52
2	Bayes	an Inference	52
	2.1	Gibbs sampling algorithm	52
	2.2	Studied samples and results	53
		2.2.1 System $(100 - x - y)NaPO_3 - xPbCl_2 - yBaCl_2 : LnF_3$ .	53
		2.2.2 System $NaPO_3$ - $ZnF_2$ - $SrF_2$ : $LnF_3$ ( $Ln = Ho$ )	56
	2.3	Discussion	58
3	Judd	Ofelt parameters prediction	58
	3.1	Dataset	58
	3.2	Experiment settings	60
	3.3	MLP structure	61
	3.4	Result	61
	3.5	Discussion	62

4	Conclusion	63
Genera	l Conclusion	64

# List of Figures

1.1	Rare Earths elements on the periodic table $[1]$	17
1.2	LS coupling scheme	21
1.3	JJ coupling scheme	21
1.4	RE energy levels degeneracy $[2]$	23
1.5	Rare Earth in atomic structure[2]	24
1.6	Energy perturbations scheme of lanthanide series atoms $[3]$	25
1.7	Energy level diagram for $\mathbb{R}^{+3}$ ions doped in a low-symmetry crystal, $LaF_3$	26
1.8	Procedure of the Judd-Ofelt analysis [4] $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	26
2.1	Venn diagram of $A$ intersection $B$	32
2.2	Non-informative and informative prior plot	37
3.1	Model of an artificial neuron	43
3.2	Sigmoid activation function and its derivative	44
3.3	Architecture of MLP	45
3.4	Overfitting	46
3.5	Underfitting	46
3.6	Robust Fit	46
3.7	applying dropout on neuron $[5]$	51
3.8	applying dropout on neural networks [5]	51
4.1	Absorption Spectra of $Er^{3+}$ doped NPE1 and NPBE1 glasses [6]	54
4.2	Absorption Spectra of $Ho^{3+}$ doped NPH1 and NPBH1 glasses [6]	54
4.3	NPH1	56
4.4	NPBH1	56
4.5	NPE1	56
4.6	NPBE1	56
4.7	NPSZH0.5	57

4.8	NPSZH1	57
4.9	NPSZH1.5	57
4.10	NPSZH2	57
4.11	Dataset head Of bulk oxide composition percentage of $Er^{+3}$ doped glasses	
	and the target JO parameters	58
4.12	dataset heat map	60

# List of Tables

1.1	Rare Earths electronic configuration table	18
1.2	Spectroscopic symbols of some $(R^{+2})$ and $(R^{+3})$ RE ions	22
3.1	Activation functions and their derivatives.	44
4.1	Compositions of glass samples [6]	54
4.2	Physical properties of glass samples [6]	54
4.3	Measured oscillator strength for $Ho^{3+}$ ions in NPH1 and NPBH1 glasses [6]	55
4.4	Measured oscillator strength for $Er^{3+}$ ions in NPE1 and NPBE1 glasse[6] .	55
4.5	Judd Ofelt parameters using least square method [7] and Bayesian inference	
	method	56
4.6	Compositions of glass samples [7]	56
4.7	Oscillator strength of $ho^{3+}$ absorption bands of NPZSHn (n=0.5, 1, 1.5,	
	$2)[7]  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  $	57
4.8	Calculated Judd Ofelt parameters using least square method [7] and	
	Bayesian inference method	58
4.9	Dataset summery	59
4.10	Summary of the options and parameters selected for our MLP model	61
4.11	The estimated Root mean squared error for training and testing data from	
	1000 iteration of MLP regression model	62

# **General Introduction**

# 1 Introduction

In 1962 B.R. Judd [8] and G.S. Ofelt [9] published separately two scientific paper on the intensities of optical transitions in rare-earth ions, we now refer their work as Judd-Ofelt theory which is the centerpiece of rare-earth spectroscopy.

Judd Ofelt theory is a mathematical model consist of calculating irradiative transition probabilities of tivalent Rare earth ions  $(RE^{+3})$ , intensity parameters  $\{\Omega_2, \Omega_4, \Omega_6\}$  are the key to calculate and predict many optical properties of Rare earth doped glasses, such as oscillator strength, luminescence branching ratios, excited-state radiative lifetimes, energy-transfer probabilities, and estimates of quantum efficiencies.

Generally the JO parameters are obtained empirically from absorption spectrum by minimizing the differences between the calculated  $f_{cal}$  and the experimental transition line strengths  $f_{exp}$  of a series of excited multiplets by standard linear fit least-squares or chi-square methods, however the computational power is cheap nowadays which hides the diverse mathematical concepts as well as the highly laborious calculations underlying the Judd Ofelt theory.

## 2 Thesis scope

As we pointed in the beginning the JO parameters are obtained by a conventional least square fit of experimental data, another alternative framework to fit the data is Bayesian inference [10], by a Bayesian linear regression we can fit the experimental data under uncertainty in more flexible way, by construct a probability distribution for the parameters of interest (JO parameters) using probability distribution of experimental data, further we will focus on use this framework to estimate the JO parameters for halogeno phosphate glasses [6, 7].

The second framework we will use along this thesis is machine learning to predict Judd Ofelt parameters, by studying the correlation between glasses composition and the JO parameters, the calculation of these parameters is considered complicated especially for the non specialist furthermore it require laboratory procedure to be done, like providing the necessary laboratory equipment, preparing the samples, the measurement of the absorption spectrum, so forth..., therefore going from calculation to prediction may avoid all these specialized procedure and enable even the non-specialist to benefit from Judd Ofelt theory framework in the innovation of glasses.

# 3 Thesis structure

In addition to this general introduction, this thesis is organized as followed :

- 1. Chapter 1 : we start by introducing Rare earth ions and their spectroscopy after that we will focus on Judd Ofelt theory and the calculation of JO intensity parameters.
- 2. Chapter 2 : The second chapter will be on using Bayesian inference framework to estimate Judd Ofelt parameters and the simulation technique that we used in our work.
- 3. Chapter 3 : we will present our machine learning method we used to predict JO parameters from glasses compositions.
- 4. Chapter 4 : we will present the different results of our experiments for each technique we used, the comparison result with related works and discus these obtained results.

In the end of this thesis, we present principals conclusions of our work, and some perspectives and future works.

# Chapter 1

# Rare Earths And Judd Ofelt Theory

## 1 Introduction

Since the beginning of the twentieth century, Rare earths (RE) have received academic and industrial attention, the study of their photophysics has given rises to intriguing science, Moreover they are widely used in technological applications such as Luminophore of TV screens, scintillators or laser materials, these are mainly made of inorganic solid materials doped with impurities (dopant ions), the optical proprieties depends on the host material which tends to be transparent, after doping with rare earths they show several fine emission bands in a spectral range from UV to near IR, depending on the dopant RE which are optically active centers.

# 2 Rare Earths

Basing on Mendeleev's periodic table of elements, Rare earths are a set of seventeen chemical elements, consists of the 15 lanthanide elements (lanthanum, cerium, praseodymium, neodymium, promethium, samarium, europium, gadolinium, terbium, dysprosium, holmium, erbium, thulium, ytterbium, and lutetium) plus Scandium and Yttrium, Sc and Y are considered RE because they have similar optical and chemical properties as the lanthanides and occur in the same ore deposits.

Rare earths are not necessarily "rare" as their name implies, Thulium and lutetium e.g. are the least abundant RE elements but they each have a greater crustal abundance than Gold (nearly 200 times greater), however they are very difficult to mine because it is unusual to find them in a high enough concentrations for economical extraction.



Figure 1.1: Rare Earths elements on the periodic table [1]

## 2.1 Rare Earths ions

RE ions are known for their magnetic and luminescence proprieties, they are very weak electronegative elements with an oxidation degree +3 ( $RE^+3$ ) (triply charged), some ions like Cerium (Cr), Praseodymium (Pr) and Terbium (Tr) have the possibility of being stabilized at degree +4 and others like Europium(Eu), Ytterbium (Y) and Samarium (Sm) at degree +2

These ions are characterized by the progressive filling of the 4f shells of their electronic configuration, the 4f shells are the main responsible of the optical proprieties of RE ions and an incomplete 4f generates intensive radiations (from UV to IR)[11]

 $[Xe]4f^n5d6s$ 

Where n = 0, 1..., 14

Elements		Ζ	А	Configuration
Scandium	$\operatorname{Sc}$	21	45	$(3d4s)^{3}$
Yttrium	Υ	39	89	$(4d5s)^{3}$
Lanthanum	La	57	139	$4f^0(5d6s)^3$
Cerium	Ce	58	140	$4f^{1}(5d6s)^{3}$
Praseodymium	$\Pr$	59	141	$4f2(5d6s)^{3}$
Neodymium	Nd	60	144	$4f^{3}(5d6s)^{3}$
Promethium	Pm	61	145	$4f^4(5d6s)^3$
Samarium	$\operatorname{Sm}$	62	150	$4f^5(5d6s)^3$
Europium	Eu	63	152	$4f^{7}(5d6s)^{2}$
Gadolinium	$\operatorname{Gd}$	64	157	$4f^{7}(5d6s)^{3}$
Terbium	$\mathrm{Tb}$	65	159	$4f^8(5d6s)^3$
Dysprosium	Dy	66	163	$4f^9(5d6s)^3$
Holmium	Ho	67	165	$4f^{10}(5d6s)^3$
Erbium	$\mathrm{Er}$	68	167	$4f^{11}(5d6s)^3$
Thulium	$\mathrm{Tm}$	69	169	$4f^{12}(5d6s)^3$
Ytterbium	Yb	70	173	$4f^{14}(5d6s)^2$
Lutetium	Lu	71	175	$4f^{14}(5d6s)^3$
				. ,

Table 1.1: Rare Earths electronic configuration table

#### 2.2 Rare Earths Spectroscopy

#### 2.2.1 Energy Levels

The energy level of triply charged ions may be obtained with high accuracy by spectral analysis or theoretical calculations, such calculations are important for the interpretation of the empirical results moreover these calculations are the only way to obtain eigenvectors [12], furthermore this method has been used to classify electronic states and evaluate energy levels of Lanthanides by Judd [8] and Actinides by Carnal with Wybourne [13]. We begin with free ion Schrodinger equation [14]:

$$\hat{H}_F \psi = E \psi \tag{1.1}$$

Where  $\psi$  is the wave function and E is the total energy of the system  $\hat{H}$  is the free ion Hamiltonian with :

$$\hat{H}_F = \hat{H}_0 + \hat{H}_1 + \hat{H}_2 \tag{1.2}$$

 $\hat{H}_0$  is the first order approximation, it presents the sum of kinetic energies of N electron and consider the outer electrons in a field produced by Xenon-like shell :

$$\hat{H}_0 = -\frac{h}{2m} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \frac{Ze^2}{r_i}$$
(1.3)

In this approximation all states of configuration have the same energy and the state of a single electron is characterized by four quantum numbers e.g.  $n, l, m_l, s$ , the degeneracy of zero order approximation is given 2(2l + 1) for all outer electrons when no equivalent electron involved however for the  $4f^N$  shell which consist of equivalent electron the degeneracy is given by the binomial coefficient  $\binom{14}{N}$  [12]

 $\hat{H}_1$  is the repulsive coulomb potential between pair of electrons i and j :

$$\hat{H}_1 = \sum_{i(1.4)$$

 $\hat{H}_2$  is the spin-orbit interaction which stands for coupling between spin angular momentum and orbital angular momentum and can be understood as magnetic dipole-dipole interaction :

$$\hat{H}_2 = \sum_{i=1}^{N} \zeta(r_i)(s_i l_i)$$
(1.5)

Where  $\zeta(r_i)$  is the spin-orbit interaction constant of the  $i^{th}$  electron, for a Coulomb with an effective nuclear charge Z':

$$\zeta(r_i) = \frac{\alpha^4 R Z'}{r_i^3} \tag{1.6}$$

With :

 $\alpha$  is the fine structure constant, R Rydberg constant, The effective charge Z' is in general different from the effective charge Z occurring in the electrostatic interactions [12].

#### 2.2.2 Schrodinger equation solution in central field approximation

The exact Schrodinger equation solutions are not possible in systems with more than one electron, but it is possible to construct a potential energy function  $U(r_i)$  which is spherically symmetric and it is a good approximation to the actual potential energy in the field of the nucleus and Xenon-like shell [14], thus (1.3) can be replaced by :

$$\hat{H}'_0 = \sum_{i=1}^N \left[ \frac{\hbar}{2m} \nabla_i^2 + U(r_i) \right]$$
(1.7)

With :

$$\sum_{i=1}^{N} U(r_i) = -\sum_{i=1}^{N} \frac{Ze^2}{r_i} + \langle \sum_{i=1}^{N} \frac{e^2}{r_{ij}} \rangle$$
(1.8)

The term  $\langle \sum_{i=1}^{N} \frac{e^2}{r_{ij}} \rangle$  stands for the average over a sphere of the electron repulsion. the  $2^{nd}$  term in (1.2) can be written as :

$$\hat{H}_{1}' = \sum_{i < j}^{N} \frac{e^{2}}{r_{ij}} - \langle \sum_{i < j}^{N} \frac{e^{2}}{r_{ij}} \rangle$$
(1.9)

The previous term along with the spin-orbit term are small enough to be treated as perturbation, thus the next step is to apply Hartree-Fock Method [15] and obtain the eigenfunction of the central field Hamiltonian :

$$\psi_{nlm_lm_s}(r,m_s) = \frac{1}{r} R_{nl}(r) Y_{ml}(\theta,\phi) \sigma(m_s)$$
(1.10)

Where  $R_{nl}$  is the radial function,  $Y_{ml}$  is the special harmonic function, and  $\sigma(m_s)$  is the spin function.

For electrons in  $4f^n$  shell we have :

$$n = 4$$
  
 $l = 3$   
 $m_l = -3, -2, ..., +3$   
 $m_s = \pm \frac{1}{2}$ 
(1.11)

Thus the wave function can be written as :

$$\psi(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\lambda_1) & \psi_2(\lambda_1) & \dots & \psi_N(\lambda_1) \\ \psi_1(\lambda_2) & \psi_2(\lambda_2) & \dots & \psi_N(\lambda_2) \\ \vdots & & & \\ \psi_1(\lambda_N) & \psi_2(\lambda_N) & \dots & \psi_N(\lambda_N) \end{vmatrix}$$
(1.12)

Where  $\psi_i(\lambda_j)$  is are spin orbitals, *i* is an electron of  $n, l, m_l, m_s$  quantum numbers and  $\lambda_j$  is the space and spin coordinates of the  $j^{th}$  electron.

The main purpose of this approximation is to use (1.12) as the basis wave functions for the perturbation terms such as the coulomb interactions in (1.9) and spin-orbit perturbation in (1.5) [14]

#### 2.2.3 LS and Intermediate Coupling

In the previous section we talked about the central field approximation in which all the electron are in spherically symetric field, the rest part of (1.2) or in more explicite form (1.5) is the spin orbit interaction which is not spherically symetric, to construct wave function in this approximation, we need to choose a coupling scheme of momentum summation.

in lighter atoms, the spin orbit interaction is not important because the coulomb interaction dominates in this case LS scheme fig(1.2) is a good approximation, whereas in heavier atoms s-o interaction becomes much stronger thus more important then JJ coupling scheme fig(1.3) would be a good choice.



Figure 1.2: LS coupling scheme



Figure 1.3: JJ coupling scheme

In our case (Rare Earths), the coulomb and s-o interactions have the same order of magnitudes thus neither LS coupling nor JJ coupling is is appropriate scheme therefore the calculation of energy levels are involved in a scheme called intermediate coupling which is developed from the LS scheme. [14]

In LS coupling we have :

$$L = \sum_{i=1}^{N} l_i$$

$$S = \sum_{i=1}^{N} s_i$$
(1.13)

And the total angular momentum :

$$J = L + S \tag{1.14}$$

The electronic state in this coupling scheme can be written as:

$$\psi = |4f^n LSJm_j\rangle \tag{1.15}$$

We use the spectroscopic symbol  ${}^{2S+1}L_j$  to name a free ion state or a multiplet, where  $L = (0, 1, ...), S = (0, \frac{1}{2}, 1, ...)$  and |L - S| < j < L + S, then we construct table(1.2):

Ζ		Element	$R^{+2}$	$R^{+3}$
57	La	Lanthanum	$4f^{1}, {}^{2}F_{\frac{5}{2}}$	$4f^{0}, {}^{1}S_{0}$
58	Ce	Cerium	$4f^2, {}^3H_4$	$4f^{1}, {}^{2}F_{\frac{5}{2}}$
59	$\Pr$	Praseodymium	$4f^{3}, {}^{4}I_{\frac{9}{2}}$	$4f^{2}, {}^{3}H_{4}^{2}$
60	Nd	Neodymium	$4f^{2}, {}^{5}I_{4}^{2}$	$4f^4,  {}^4I_{\frac{9}{2}}$
67	Но	Holmium	$4f^{11}, {}^{4}I_{\frac{15}{2}}$	$4f^{11},  {}^5I_8$
68	Er	Erbium	$4f^{12}, {}^3H_6$	$4f^{11},  {}^{4}I_{\frac{15}{2}}$
				2

Table 1.2: Spectroscopic symbols of some  $(R^{+2})$  and  $(R^{+3})$  RE ions

Since there are states with the same L and S quantum numbers, a new quantum number  $\tau$  (the seniority number) has been added to differentiate between these states, thus the eigenfunctions in this intermediate coupling scheme are written as :

$$\psi(nlJ) = \sum_{\tau LS} a_{\tau LSJ} \left| 4f^n \tau LSJ \right\rangle \tag{1.16}$$

With :

$$a_{\tau LSJ} = \sum_{\tau' L'S'} \langle 4f^n \tau LSJ | \hat{H}_1 + \hat{H}_2 | 4f^n \tau' L'S'J' \rangle$$
(1.17)

this new eigenstate describes the energy state of the Hamiltonian including Coulomb and spin-orbit interaction [8, 14]

In the below Diagram representing the interactions that leads to the splitting of the electronic energy levels for a lanthanide ion fig(1.4) [2]



Figure 1.4: RE energy levels degeneracy [2]

#### 2.2.4 Crystal field approximation

In Crystal Field Theory (CFT), the lanthanides are considered to be in crystalline host (fig1.5) thus the RE are perturbed by a crystal field which decreases the symmetry of the spherically-symmetric free ion Hamiltonian.

$$\hat{H} = \hat{H}_f + \hat{H}_{cf} \tag{1.18}$$



Figure 1.5: Rare Earth in atomic structure<sup>[2]</sup>

The crystal field interactions Hamiltonian are written as sum of spherical tensor operators :

$$\hat{H}_{cf} = \sum_{k,q} A_{k,q} \sum_{i} r_i^k Y_{kq}(\theta, \phi)$$
(1.19)

Where  $Y_{kq}(\theta, \phi)$  are are spherical harmonic functions and (k, q) are combinations that give non-zero matrix elements depends on the symmetry of RE sites.

$$A_{k,q} = -e \sum_{i} \frac{Z_i Y_{kq}(\theta, \phi)}{R_i^{k+1}}$$
(1.20)

 $R_i$  are the positions of the surrounding atoms composing the crystal.

These terms are considered in the calculation of the energy levels and they are estimated empirically from experimental data [2].

the effect of the crystal field interaction are weaker than coulomb interactions or spin orbit interaction (fig1.6) [11, 16]



Figure 1.6: Energy perturbations scheme of lanthanide series atoms [3]

The degenerancy in electric dipole approximation  $g = 2j + \frac{1}{2}$  whereas g = 2j + 1 in magnetic dipole approximation

energy levels of  $R^{+3}$  of lanthanides based on computed crystal field energies in the range  $0 - 50000 cm^{-1}$  are depicted in (fig1.7)[17]



Figure 1.7: Energy level diagram for  $R^{+3}$  ions doped in a low-symmetry crystal,  $LaF_3$ 

# 3 Judd Ofelt theory

Judd Ofelt (JO) theory [8, 9] was developed to calculate the radiation transition probabilities between the 4f energy levels of excited trivalent rare earth ions within a material, it is based on the free-ion and single configuration approximations explained in the previous sections.

This section represents an essential summary and some important formulas for the JO theory and the procedure for the Judd-Ofelt analysis as shown in fig1.8



Figure 1.8: Procedure of the Judd-Ofelt analysis<sup>[4]</sup>

For a free ion, electrical dipole transitions between configurations of the same parity are prohibited by Laporte's parity rule, however they are observed between 4f shells in RE thus they are forced transitions, the main reason is that RE are in an asymmetric crystal field.[18]

Such transition are imposed by ED selection rules [19] :

$$\Delta l = \pm 1$$
  

$$\Delta S = 0$$
  

$$\Delta L \le 2l$$
  

$$\Delta J \le 2l$$
  
(1.21)

Whereas in magnetic dipole approximation :

$$\Delta L = \pm 0$$
  

$$\Delta S = 0$$
(1.22)  

$$\Delta J = 0, \pm 1, 0 \not\rightarrow 0$$

These transition are electrons decay from excited state  $|4f^n LSJ\rangle$  to lower energy final state  $|4f^n L'S'J'\rangle$ , the dacay lifetime is given by :

$$\tau = \frac{1}{\sum_{L'S'J'} (A_{LSJ \to L'S'J'})}$$
(1.23)

Where  $A_{SLJ\to L'S'J'}$  is the total spontaneous emission probability, it takes into account the ED and MD transition:

$$A_{LSJ \to L'S'J'} = A_{LSJ \to L'S'J'}^{ED} + A_{LSJ \to L'S'J'}^{MD}$$
(1.24)

#### 3.1 Electric dipole approximation

The spontaneous emission probability  $A_{LSJ\to L'S'J'}^{ED}$  is related to the oscillator strength  $f^{ED}$  and given by:

$$A_{LSJ\to L'S'J'}^{ED} = \frac{2\pi e^2}{m_e \epsilon_0 \lambda^2 c} f_{ED}^{emi}$$
(1.25)

Where  $\lambda$  is the mean wavelength :

$$\lambda_{j \to j'} = \frac{\int_{j \to j'} \lambda I(\lambda) d\lambda}{\int_{j \to j'} I(\lambda) d\lambda}$$
(1.26)

The ED oscillator strength  $f_{ED}^{emi}$  given by (1.27) is obtained by solving the electric dipole matrix elements  $\langle \psi | \vec{P} | \psi' \rangle$ ,  $\vec{P}$  is the ED operator.

$$f_{ED}^{emi} = \frac{8\pi m_e c\nu\chi}{3h(2J+1)} \sum_{k=2,4,6} \Omega_k \mid \langle 4f^n SLJ \mid U^{(k)} \mid 4f^n S'L'J' \rangle \mid^2$$
(1.27)

The JO theory succeed in finding an expression for the ED absorption intensity, by proposing four assumptions in order to simplify the calculations [20]:

- 1. All sub-levels of  $4f^{n-1}5d$  are degenerate in J.
- 2. the  $4f \to 4f$  and  $4f \to 4f^{n-1}5d$  energy separation are the same.
- 3. All Stark sub-levels within the ground state are equally populated
- 4. The host material is optically isotropic

Where :

 $\chi = \frac{(n^2+1)^2}{9n}$  is the local-field correction factor,  $\Omega_k$  are JO parameters with k = 2, 4, 6 arises from selection rule simplification  $\Delta J = 2, 4, 6$  and the parameterization of the crystal field [21], furthermore  $\Omega_2$  is related to the co-valency of the chemical bond between the RE ions and the ligand ions thus it is more sensitive to the local symmetry of the environment while  $\Omega_4$  and  $\Omega_6$  less sensitive [22], Moreover  $\Omega_4$  and  $\Omega_6$  may relate to the viscosity and rigidity of the environment.

The dimensionless doubly-reduced matrix elements  $|\langle U^{(k)} \rangle|$  are values already tabulated in [23]

## 3.2 Magnetic dipole approximation

the the oscillator strength of magnetic dipole transitions only depends on one reduced matrix element  $\langle 4f^n LSJ | L + gS | 4f^n L'S'J' \rangle$ [20]:

$$f_{MD}^{emi} = \frac{hn^3}{6m_e c\lambda(2J+1)} \left\langle 4f^n LSJ \right| L + gS \left| 4f^n L'S'J' \right\rangle$$
(1.28)

Where L + gS is the magnetic dipole operator and g = 2.002319304362 is the electron factor [22], however this term is generally negligible in compare of Electric dipole term.

#### **3.3** Judd-Ofelt parameters calculation

The calculation of JO parameters depends on absorption coefficient  $\alpha(\lambda)$  which is part of the experimental line strength of the transition:

$$f^{exp} = \frac{3hc(2J+1)n}{8\pi^3\lambda e^2 N_a \chi} \int \alpha(\lambda)$$
(1.29)

Next we compare (1.29) with the JO theoretical line strength calculation to fit JO parameters .

$$f^{cal} = e^2 \sum_{k=2,4,6} \Omega_k \mid \langle 4f^n SLJ \mid U^{(k)} \mid 4f^n S'L'J' \rangle \mid^2$$
(1.30)

thus :

$$\begin{bmatrix} f_1^{exp} \\ f_2^{exp} \\ \vdots \\ f_N^{exp} \end{bmatrix} = \begin{bmatrix} |\langle U_1^{(2)} \rangle|^2 & |\langle U_1^{(4)} \rangle|^2 & \langle U_1^{(4)} \rangle|^2 \\ |\langle U_2^{(2)} \rangle|^2 & |\langle U_2^{(4)} \rangle|^2 & \langle U_2^{(4)} \rangle|^2 \\ \vdots & \vdots & \vdots \\ |\langle U_N^{(2)} \rangle|^2 & |\langle U_N^{(4)} \rangle|^2 & \langle U_N^{(4)} \rangle|^2 \end{bmatrix} \begin{bmatrix} \Omega_2 \\ \Omega_4 \\ \Omega_6 \end{bmatrix}$$
(1.31)

Hence :

$$y = x\Omega \to \Omega = (x^T x)^{-1} x^t y \tag{1.32}$$

Where y is vector of  $f^{exp}$  and  $x^T$  is the transpose matrix of  $(U^{(k)})^2(x)$  and  $\Omega$  is vector of JO parameters.

After we obtain the JO parameters, we can use it to calculate the theoretical line strength in (1.30) then measure the root mean squared error :

$$RMSE = \left[\frac{(f^{exp} - f^{cal})^2}{N - 3}\right]^{\frac{1}{2}}$$
(1.33)

Where N is the number of transitions and 3 is the number of JO parameters [21].

#### 3.4 Conclusion

In this chapter we have outlined the required background in the context of our work : Rare earth ions, its spectroscopy and a brief introduction to Judd Ofelt theory with emphasis on Judd Ofelt intensity parameters We also present the classical approach to obtain JO parameters using least square method however the next chapters we will introduce another approaches in determining and predicting these parameters

# Chapter 2

# **Bayesian Inference**

## 1 Introduction

In statistics, there are two major paradigms to inference, conventional (or frequentist) and Bayesian, the Bayesian methods provide a consistent framework for inference and making decisions under uncertainty, unlike frequentist paradigm, Bayesian inference gave us the possibility to incorporate scientific hypothesis by using prior distribution in our problem of interest, and fitting a probability model to a set of data using probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations [24, 25].

In this chapter we will discover Bayesian approach to inference principals and use this framework in determining the JO parameters.

# 2 Conditional Probability

Conditional probability is the probability of one event occurring with some relationship to one or more other events, and we note:

$$P(A \mid B) = \frac{P(A \cap B)}{P(A)}$$
(2.1)

where  $P(A \mid B)$  is the event A given B to occur,  $P(A \cap B)$  the intersection of all A and B set of events as shown in the below figure.



Figure 2.1: Venn diagram of A intersection B

The conditional probability thus measures the probability of B given that we know that A has occurred. The numerator in (2.1) is the probability that both of them occur, and the denominator rescales this number in order for conditional probabilities to satisfy the Kolmogorov axioms [26, 27]

# 3 Bayes theorem

In the previous section we introduced conditional probability the next steps in deriving Bayes formula are as follow.

From (2.1) we have :

$$P(A \cap B) = P(B \mid A)P(A) \tag{2.2}$$

Also :

$$P(B \cap A) = P(A \mid B)P(B) \tag{2.3}$$

Hence from  $P(A \cap B) = P(B \cap A)$ :

$$P(B \mid A)P(A) = P(A \mid B)P(B) \to P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$
(2.4)

This equation known as Bayes theorem and is the basis of our statistical inference.

If we have a partition S of  $B_i$  sub-events then for any  $A \subset S$ :

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$
(2.5)

Then (2.4) becomes :

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$
(2.6)

Suppose we have continuous parameter  $\theta$  with density  $f(\theta)$  in range [a, b] and discrete random data x, in this case, the total probability of x is :

$$P(x) = \int_{a}^{b} P(x \mid \theta) P(\theta) d\theta$$
(2.7)

And Bayes Formula becomes:

$$P(\theta \mid x) = \frac{P(x \mid \theta)P(\theta)}{\int_{a}^{b} P(x \mid \theta)P(\theta)d\theta}$$
(2.8)

We call  $P(\theta \mid x)$  a posterior,  $P(\theta)$  a prior,  $P(x \mid \theta)$  the likelihood and the denominator presents a normalized constant, hence:

$$posterior = \frac{likelihood \times prior}{constant}$$
(2.9)

When we apply Bayes rule, we usually wish to infer the parameters  $\theta$  and the data x is already given, the normalized constant in (2.7) is intractable and hard to calculate most of the time.

Since we approximate our calculations by drawing samples from the posterior we don't include the normalized factor and re-write Bayes rule in term of proportionality [26, 28, 29]

$$posterior \propto likelihood \times prior \tag{2.10}$$

## 4 Linear regression

Before presenting Bayesian viewpoint of linear regression, we introduce conventional linear regression in the next section.

#### 4.1 Conventional linear regression

Consider our problem of interest, which is determining the JO parameters, we had from (1.27):

$$f_{ED}^{emi} = \frac{8\pi m_e c\nu\chi}{3h(2J+1)} \sum_{k=2,4,6} \Omega_k \mid \langle 4f^n SLJ \mid U^{(k)} \mid 4f^n S'L'J' \rangle \mid^2$$

The above equation can be reformulated as follows [10]:

$$\frac{3h(2(j+1))}{8\pi^2 m c \nu \chi} f_i = \Omega_2 (U_i^2)^2 + \Omega_4 (U_i^4)^2 + \Omega_6 (U_i^6)^2$$
(2.11)

Where  $f_i = S_{ED}^{exp}$ , hence we can re-describe the model by:

$$y_i = \Omega_2 x_i^{(2)} + \Omega_4 x_i^{(4)} + \Omega_6 x_i^{(6)} + \epsilon_i$$
(2.12)

Where, the response variable  $y_i = \frac{3h(2(j+1))}{8\pi^2 m c \nu \chi} f_i$  with i = 1, 2, 3, ..., n number of observations,  $\Omega_2, \Omega_4, \Omega_6$  are our parameters of interest,  $x_i^{(j)} = (U_i^{(j)})^2$  is the  $j^{th}$  with j = 2, 4, 6 explanatory variable of the  $i^{th}$  observation,  $\epsilon_i$  is the  $i^{th}$  associated error.

Thus we can establish our model in matrix form as follows :

$$Y = \Omega X + \epsilon \tag{2.13}$$

Where Y is  $n \times 1$  vector of the response variable,  $\Omega = (\Omega_2, \Omega_4, \Omega_6)$ , X is  $n \times 3$  matrix and  $\epsilon$  is  $n \times 1$  residuals vector, furthermore, by Central limit theorem the residuals  $\epsilon$  are assumed to be normally distributed (Gaussian) with mean 0 and standard deviation  $\sigma^2$ , thus ( $\epsilon \sim N(0, \sigma^2)$ )

We assume that the probability density function of the  $i^{th}$  observation  $y_i$  is an normal distribution with mean  $\Omega_j x_i^{(j)}$  and variance  $\sigma^2$ :

$$y_i \sim N(\Omega_j x_i^{(j)}, \sigma^2)$$

We can now obtain the likelihood function by :

$$L(Y \mid \Omega, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-(y_i - \Omega x_i)^2}{2\sigma^2}\right)$$
$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(\frac{-(Y - \Omega X)^T(Y - \Omega X)}{2\sigma^2}\right)$$
(2.14)

In order to make inference, we can obtain  $\Omega$  by maximizing the likelihood or its log for mathematical convenience, which is given by:

$$Log(L(Y \mid \Omega, \sigma^{2})) = -\frac{n}{2}Log(2\pi) - \frac{n}{2}Log(\sigma^{2}) - \frac{1}{2\sigma^{2}}(Y - \Omega X)^{T}(Y - \Omega X)$$
(2.15)

Differentiating equation (2.15) with respect to  $\Omega$  and solving for  $\Omega$ , we have :

$$\frac{\partial Log(L(Y \mid \Omega, \sigma^2))}{\partial \Omega} = \frac{1}{2\sigma^2} 2X^T Y - 2X^T X \Omega = 0$$

$$\Omega = (X^T X)^{-1} X^T Y$$
(2.16)

Which is the same result in (1.32), furthermore the variance of  $\Omega$  is :

$$Var[\Omega] = Var[(X^{T}X)^{-1}X^{T}Y]$$
  
=  $(X^{T}X)^{-1}XVar[y]((X^{T}X)^{-1}X)^{T}$   
=  $\sigma^{2}(X^{T}X)^{-1}X^{T}X(X^{T}X)^{-1}$   
=  $\sigma^{2}(X^{T}X)^{-1}$  (2.17)

Similarly the maximum likelihood estimation for  $\sigma^2$  is :

$$\frac{\partial Log(L(Y \mid \Omega, \sigma^2))}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(Y - \Omega X)^T (Y - \Omega X)}{2(\sigma^2)^2} = 0$$
$$\sigma^2 = \frac{(Y - \Omega X)^T (Y - \Omega X)}{n}$$
(2.18)

#### 4.2 Bayesian linear regression

In Bayesian linear regression the parameters  $\Omega$  and  $\sigma^2$  can be estimated in different ways, usually through stochastic simulation methods such as Markov chain monte carlo (MCMC), from classical viewpoint and by the theorem of central limit the asymptotic distribution of our parameter of interest will be  $\Omega \sim N((X^T X)^{-1} X^T Y, \sigma^2 (X^T X)^{-1})$  [30, 31], in other hand the choice of  $\Omega$  distribution may differ according to the choice of the conjugate priors.

In this section we will present Bayesian linear model to estimate Judd Ofelt parameters, we will use the same perspective established in (2.12) and (2.13), details in the below Consider our model :

$$Y = \Omega X + \epsilon$$

With Y is  $n \times 1$  vector,  $\Omega = (\Omega_2, \Omega_4, \Omega_6)$  and X is  $n \times 3$  matrix.

In the Bayesian viewpoint, we formulate linear regression using probability distributions rather than point estimates. The response Y is not estimated as a single value, but is assumed to be drawn from a probability distribution ,the model for Bayesian Linear Regression with the response Y sampled from a normal distribution depicted in the previous section (2.14).

$$L(Y \mid \Omega, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(\frac{-(Y - \Omega X)^T (Y - \Omega X)}{2\sigma^2}\right)$$

#### 4.2.1 Prior

Since we don't have any prior knowledge about our data we have an alternative way to define a prior, non-informative priors or flat [10]



Figure 2.2: Non-informative and informative prior plot

Box [32] and Tiao [33, 34] define a non-informative prior as a prior which provides little information relative to the experiment, Bernardo [35] and Smith [36] use a similar definition they say that non-informative priors have minimal effect relative to the data. Jeffreys also described how to construct such a prior for a multiparameter model( $\Omega, \sigma^2$ ), which is based on the Fisher information function[37, 38]

In our model where  $y_i \sim N(\Omega x_i, \sigma^2)$  for single observation, the Jeffreys proposal of a non-informative prior pdf is :

$$p(\Omega, \sigma^2) \propto \sqrt{det(FI(\Omega, \sigma^2))}$$
 (2.19)

The fisher information matrix element is defined by :

$$FI(\Omega, \sigma^2) = -E\left[\frac{\partial^2 Log(L(Y \mid \Omega, \sigma^2))}{\partial \theta_i \partial \theta_j}\right]$$
(2.20)

Where  $\{\theta_i, \theta_j\} \in \{\Omega, \sigma^2\}$ 

Thus :

$$FI(\Omega, \sigma^{2}) = \begin{pmatrix} \frac{\partial^{2}Log(L(Y|\Omega, \sigma^{2}))}{\partial\Omega^{2}} & \frac{\partial^{2}Log(L(Y|\Omega, \sigma^{2}))}{\partial\Omega\partial\sigma^{2}} \\ \frac{\partial^{2}Log(L(Y|\Omega, \sigma^{2}))}{\partial\sigma^{2}\partial\Omega} & \frac{\partial^{2}Log(L(Y|\Omega, \sigma^{2}))}{\partial(\sigma^{2})^{2}} \end{pmatrix}$$

$$= -E \begin{pmatrix} \frac{1}{\sigma^{2}} & \frac{2(Y - \Omega X)}{\sigma^{2}} \\ \frac{2(Y - \Omega X)}{\sigma^{2}} & \frac{2(Y - \Omega X)^{2}}{\sigma^{4}} - \frac{1}{\sigma^{2}} \end{pmatrix}$$
(2.21)

Since  $E[Y - \Omega X] = 0$  and  $E[(Y - \Omega X)^2] = \sigma^2$ , then :

$$FI(\Omega, \sigma^2) = \begin{pmatrix} -\frac{1}{\sigma^2} & 0\\ 0 & \frac{1}{\sigma^2} \end{pmatrix}$$
(2.22)

Next, we construct Jeffreys prior :

$$p(\Omega, \sigma^2) \propto \begin{vmatrix} -\frac{1}{\sigma^2} & 0\\ 0 & \frac{1}{\sigma^2} \end{vmatrix}^{\frac{1}{2}} \propto \frac{1}{\sigma^2}$$
(2.23)

Which is a very popular and a default choice of a non-informative prior for such a model.

#### 4.2.2 Posterior

From (2.10) we have :

$$P(\Omega, \sigma^2 \mid Y) \propto L(Y \mid \Omega, \sigma^2) \times p(\Omega, \sigma^2)$$
(2.24)

Thus :

$$P(\Omega, \sigma^2 \mid Y) \propto \frac{1}{\sigma^2} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(\frac{-(Y - \Omega X)^T (Y - \Omega X)}{2\sigma^2}\right)$$
(2.25)

This posterior differs from the likelihood function only in the leading exponent, the absolute value of the exponent for  $\sigma^2$  is increased from n/2 to n/2+ 1, which is an asymptotically irrelevant modification of the likelihood function. this result highlights that with large samples, the prior may matter very little in affecting posterior inference.

#### 4.2.3 Markov Chain Monte Carlo simulation

In order to estimate the model parameters in Bayesian framework, Markov Chain Monte Carlo methods are useful tool to sample from the posterior, the term Markov chain refers to a discrete-time stochastic process on a general state space that has the Markov property: "the future is independent of the past given the present state" [39, 40].

We can summarize Markov chain process in mathematical expression shown below :

$$P(\Omega_{t+1}, \sigma^2 \mid y_1, y_2, \dots, y_n) = P(\Omega_{t+1}, \sigma^2 \mid y_t)$$
(2.26)

There are two major MCMC algorithms : Gibbs sampling and Metropolis Hastings, in this section we briefly introduce Gibbs sampling since, Metropolis Hastings is beyond the scope of our work.

The Gibbs Sampling is a stochastic simulation method via Markov chain, applied when the joint posterior distribution has no famous form (normal, gamma...) but the full conditionals do, the main advantage of this algorithm is the covergence of the Markov chain is ensured without any tuning, however, Gibbs sampling requires fully conditional probability distributions to be specified for all the parameters  $(\Omega, \sigma^2)$  [31, 41]

To derive the full conditionals we have :

$$P(\Omega \mid Y, \sigma^2) = \frac{P(\Omega, \sigma^2 \mid Y)}{P(\sigma^2 \mid Y)}$$
(2.27)

$$P(\sigma^2 \mid Y, \Omega) = \frac{P(\Omega, \sigma^2 \mid Y)}{P(\Omega \mid Y)}$$
(2.28)

From (2.27) and since  $P(\sigma^2 \mid Y) = \frac{1}{\sigma^2}$  we get :

$$P(\Omega \mid Y, \sigma^{2}) = \frac{P(\Omega, \sigma^{2} \mid Y)}{\sigma^{-2}}$$

$$\propto \frac{1}{(2\pi\sigma^{2})^{\frac{n}{2}}} exp\left(\frac{-(\Omega X - Y)^{T}(\Omega X - Y)}{2\sigma^{2}}\right)$$

$$\propto \frac{1}{(2\pi\sigma^{2})^{\frac{n}{2}}} exp\left(\frac{-(X^{T}X)^{2}(\Omega - (X^{T}X)^{-1}X^{T}Y)^{T}(\Omega - (X^{T}X)^{-1}X^{T}Y)}{2\sigma^{2}}\right)$$

$$\propto \frac{1}{(2\pi\sigma^{2})^{\frac{n}{2}}} exp\left(\frac{-(\Omega - (X^{T}X)^{-1}X^{T}Y)^{T}(\Omega - (X^{T}X)^{-1}X^{T}Y)}{2(\sigma(X^{T}X)^{-1})^{2}}\right)$$
(2.29)

Hence the full conditional for  $\Omega$  is :

$$P(\Omega \mid Y, \sigma^2) \sim N((X^T X)^{-1} X^T Y, \sigma(X^T X)^{-1})$$
(2.30)

From (2.28) , with  $P(\Omega \mid Y) \propto 1$ , we derive  $P(\sigma^2 \mid Y)$  as following:

$$p(\sigma^{2} \mid Y, \Omega) \propto P(\Omega, \sigma^{2} \mid Y)$$

$$\propto \frac{1}{\sigma^{2}} \frac{1}{(2\pi\sigma^{2})^{\frac{n}{2}}} exp\left(\frac{-(Y - \Omega X)^{T}(Y - \Omega X)}{2\sigma^{2}}\right)$$

$$\propto \frac{1}{(\sigma^{2})^{1 + \frac{n}{2}}} \frac{1}{(2\pi)^{\frac{n}{2}}} exp\left(\frac{-(Y - \Omega X)^{T}(Y - \Omega X)}{2\sigma^{2}}\right)$$
(2.31)

This full conditional is easily seen to be an inverse gamma distribution [10, 42] with parameters  $\alpha = \frac{n}{2}$  and  $\beta = \frac{(Y - \Omega X)^T (Y - \Omega X)}{2}$ , thus :

$$P(\sigma^2 \mid Y, \Omega) \sim IG(\frac{n}{2}, \frac{(Y - \Omega X)^T (Y - \Omega X)}{2})$$
(2.32)

With the conditionals derived, we can implement Gibbs sampling from the full conditionals by [42]:

- 1. Establishing starting values for the parameters
- 2. Sampling  $\Omega$  from its multivariate normal distribution with  $\sigma^2$  fixed
- 3. Sampling  $\sigma^2$  from its inverse gamma distribution with  $\Omega$  fixed

# 5 Conclusion

In this chapter we outlined a Bayesian framework to estimate JO intensity parameters by sampling from the joint posterior distribution  $(P(\Omega, \sigma^2 | Y))$ , we put more attention in deriving the full conditionals to apply MCMC Gibbs sampling method, we will use the framework we constructed to estimate our parameter of interest from some Rare earth doped glasses.

# Chapter 3

# Judd Ofelt parameters prediction

## 1 Introduction

In recent decades, the world has experienced a real explosion in the volume of data which is the main reason that made scientists use data to infer and obtain information about many unexplained phenomena therefore a smart data analysis leads to a significant scientific progress, one of the domains that deals with data is machine learning.

Machine learning is a sub-domain of intelligence artificial focuses on the development of models capable of representing certain characteristics, learn and detect some statistical pattern from data in order to accomplish various tasks, the term intelligence stands for the ability of these models to generalize, i.e. to extract information from the studied data during an updating process called training, and use these information to automatically infer another information from new data.

there are many predictive technique in machine learning, we will focus on supervised regression model to solve our task.

#### 1.1 Regression

Regression is to map an input data to a numerical value [29]. In another words for an input  $X_i \in \mathbb{R}^d$  that represents d dimensional features vector and a continues output space  $Y \subset \mathbb{R}$ , the learning algorithm is asked to produce a function  $f : \mathbb{R}^d \to \mathbb{R}^n$  that maps any given input  $X_i$  to a corresponding value  $y \in Y$ . Examples Neural Networks, Support Vector Regression, Linear Regression, Polynomial Regression ...

#### **1.2** States of the art

In order to predict JO parameters M.Konstantinidis et al[43] used the composition percentage to investigate the relationship between JO parameters and the bulk composition percentage of  $Er^{+3}$  doped tellurite glasses using Support vector machine regression (or SVR) with sparse Principal Component Analysis (sparse PCA) in data prepossessing for more details see [43, 44, 45, 46, 47], the collected data is of 135 observation and 25 glasses oxide composition thus they composed a matrix of  $135 \times 25$ 

However we will use a similar technique in machine learning regression, a Multylayer perceptions to fit the data and predict the JO parameters from the same dataset of glasses oxide composition, we will give, in this chapter, a global principals about artificial neural networks after that we will focalize on deep neural network (DNN) mainly perceptrons and our technique Multilayer perceptron (MLP)

## 2 Neural networks and deep learning

The idea goes back to 1943 by neurophysiologist Warren McCulloch and mathematician Walter Pitts, they made a model of connected circuits in order to simulate intelligent behavior [29, 48], they called it "threshold logic" which basically converts continues input to discreet output.

Deep learning is a subset of machine learning methods that deals with artificial neural networks (ANN). A neural network is a mathematical function that can be used in a different machine learning tasks such as classification, regression, clustering, feature extraction and dimensional reduction ... etc, All neural networks architectures consist of stack interconnected layers, each layer contains a number of neurons, the simplest neural network has an input layer and output layer as for deep neural networks (DNN) they contains a multiple hidden layers (the intermediate layers between the input and output)[29]. The depth of the network allows it to learn more and more representation and patterns about the inputs data, thus allowing them to achieve a higher accuracy especially in none-linear problems.

There are a lot of DNN architectures each has its own advantages at some tasks like convolution neural network (CNN), Recurrent neural networks (RNN)...etc, in our case we will use multi-layer preceptron (MLP).

## 2.1 Perceptron(artificial neuron)

The perceptron is the building block of MLP.It compute dot product of its weights and the inputs followed by an activation function. A single perceptron can perform a simple classification or regression task depending on the used activation function, the output of a percepton is calculated directly as follows.

Consider a single neuron with inputs  $\vec{x} = (x_1, x_2, \dots, x_n)$  with weights  $\omega_1, \omega_2, \dots, \omega_n$  and a bias b as illustrated in figure 3.3



Figure 3.1: Model of an artificial neuron

the potential of the neuron can be computed as :

$$n = \sum_{i=1}^{n} \omega_i x_i + b \tag{3.1}$$

let's consider a well known activation function Sigmoid :

$$f(n) = \frac{1}{1 + e^{-n}} \tag{3.2}$$

The output of the perceptron y is :

$$y = f(n) = f\left(\sum_{i=1}^{n} \omega_i x_i + b\right)$$
(3.3)

In the following table we present some of the most used activation functions:

Name	Function	Derivative
Sigmoid	$\phi(x) = \frac{1}{1 + e^{-x}}$	$\phi'(x) = \phi(x)(1 - \phi(x))$
TanH	$\phi(x) = \frac{2}{1 + e^{-2x}} - 1$	$\phi'(x) = 1 - \phi(x)^2$
ReLU	$\phi(x) = \begin{cases} 0 & x \le 0\\ x & x > 0 \end{cases}$	$\phi'(x) = \begin{cases} 0 & x \le 0\\ 1 & x > 0 \end{cases}$
Leaky ReLU	$\phi(x) = \begin{cases} \alpha x & x \le 0\\ x & x > 0 \end{cases}$	$\phi'(x) = \begin{cases} \alpha & x \le 0\\ 1 & x > 0 \end{cases}$

Table 3.1: Activation functions and their derivatives.

An activation function must be none-linear, continious and non-polynomial.[29, 49, 50] figure (3.5) presents a plot of the most commonly used activation function Sigmoid.



Figure 3.2: Sigmoid activation function and its derivative

However a single perceptron are trivial in complex representations, thus we present Multilayer perceptron (MLP) in the flowing section which is more flexible and universal approximator [29, 49]

## 2.2 Multilayer perceptron

The Multilayer perceptron is a feed-forward network consist of multiple fully connected layers of perceptrons (every perceptron in a layer is connected to all perceptron in the next layer), the purpose behind using MLP is to solve none-linear (complex) tasks. The architecture of our MLP model illustrated in figure 3.6 of 25 input and 3 output which are the JO parameters { $\Omega_2, \Omega_4, \Omega_6$ }, and of a given number of hidden layers



Figure 3.3: Architecture of MLP

The output of the model is computed sequentially, layer by layer, it starts by assuming  $\vec{y}_0 = \vec{x}$  then it assign that  $\vec{x}_i = \vec{y}_{i-1}$  for the hidden layer  $H_i$ , the activation function and the weights are given by the network, therefore, the output of the current layer depends only on the output of the previous layer and the final output of interest is  $\vec{y}_k = \{\Omega_2, \Omega_4, \Omega_6\}$  produced from the output layer.

# 3 Training Neural neural networks

In order to achieve the desired results using an ANN it needs to be trained. Fitting the network is to find optimal parameters (wights, bias ...) for solving our task, in the beginning of the training process the parameters needs to be initialized.

Learning process is then feeding the training data to the model in iterative way, where the output produced each iteration are compared to the actual values and the network is repeatedly adjusted till it returns the desirable result, this technique is the so called supervised learning, unsupervised learning will not be discussed here.

Some of the problems that we may encountered in this process are underfitting and overfitting.

1. Underfitting : Is when a machine learning model can not properly learn from the training data (have low accuracy) [29]. Some of the reasons why underfitting happens in neural networks is to have a small model or using a linear model with none linear dataset (features in the dataset are complex). Another reason is the noisy data (containing wrong labels).

2. Overfitting : is when a machine learning model gives a high prediction accuracy on the training data, but the prediction accuracy gets low if the model tested on previously unseen data (a data that was not present during the training) [29], another term for describing overfitting is "high generalization error". Overfitting occurs when the model gets closely fit to the training data, this is because the training data is not all the possibilities of input data. A good model should have a good accuracy on the training data and the other. In other word, ( it should be able to generalize).

The main reason for overfitting is the learning process performed for too long (many epochs). Also using a big model (too much layers) for small dataset.

The elustrated figures (3.7, 3.8, 3.9) show the difference between overfitting, underfitting and Robust fit



Figure 3.4: Overfitting

Figure 3.5: Underfitting

Figure 3.6: Robust Fit

To avoid the previous problems, the data needs to be divided into training and testing sets, it is preferable to use 80% of the available data as training set and the rest as testing set [51]

#### 3.1 Loss function

In training process the network is meant to be adjusted to return the desirable output each epoch, this adjustment is based on the calculation and minimizing the error (Loss) There exists many metrics to calculate the error, we will focalize on mean squared error (MSE).

Consider a training set of  $\{(x_i, \Omega_i)\}$  where x is glasses composition and  $\Omega$  JO parameters, every x from the training set yields an output y and we want y be as close as possible to  $\Omega$ , we define the error as :

$$e_i = y_i - \Omega_i \tag{3.4}$$

Then the mean squared error for is :

$$E_p = \frac{1}{m} \sum_{i=1}^{m} (y_i - \Omega_i)^2$$
(3.5)

Where m is the number of output neurons (3 in our case).

#### 3.2 Gradient decent

Gradient descent is an optimization algorithm for finding a local or global minimum for a differential function.

$$minf(\omega) \quad and/or \quad \omega^* = arg(minf(\omega))$$
 (3.6)

This is effective when f is convex, the algorithm is iterative and aims to find the the direction where f can decrease, so it may not give the exact value of  $\omega^*$ , but it can reach a closer value, for mathematical convenience we introduce the learning rate  $0 < \eta < 0$ , it controls how fast the gradient descent algorithm works in order to reach the global minimum.

The general form of gradient descent formula :

$$\omega_{k+1} = \omega_k - \eta \nabla f(\omega_k) \tag{3.7}$$

We give the gradient descent algorithm :

Algorithm 1 Gradient Descent Require: Differentiable loss function fRequire: Maximum number of iterations TInitialize  $\omega = \omega_0$ for  $t = 1 \rightarrow T$  do  $\omega_{k+1} = \omega_k - \eta \nabla f(\omega_k)$ EndFor Return  $\omega_k$ 

Note: Generally the gradient descent stops when it reach the global minimum  $\omega^*$  which is hard to guarantee, an alternative way to implement a stoping condition is to run the algorithm for fixed number of steps in order to reach a asymptotic convergence to the minimum.

## 3.3 Backpropagation

To adjust a DNN, we need to minimize the loss function by applying the gradient descent in all its layers, suppose we aim to minimize mean squared error to adjust the weights between  $i^{th}$  and  $j^{th}$  layer.

By applying the gradient descent we calculate the adjustment as follows :

$$\Delta\omega_{ij} = -\eta\nabla E = -\eta \frac{\partial E}{\partial\omega_{ij}} \tag{3.8}$$

By chain rule we get :

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial \omega_{ij}} \tag{3.9}$$

from (3.1) and (3.3) we have:

$$\frac{\partial E}{\partial y_j} = y_i - \Omega_i \tag{3.10}$$

$$\frac{\partial y_j}{\partial \omega_{ij}} = \frac{\partial y_i}{\partial n} \frac{\partial n}{\partial \omega_{ij}} = f'(n)y_i \tag{3.11}$$

Thus :

$$\Delta\omega_{ij} = -\eta(y_i - \Omega_i)f'(n)y_i \tag{3.12}$$

For mathematical convenience let :

$$\delta_j = \frac{\partial E}{\partial n_j} = (y_i - \Omega_i) f'(n) \tag{3.13}$$

Then :

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial\omega_{ij}} = -\eta \delta_j y_i \tag{3.14}$$

However this result is applicable only in the output layer, for the hidden layers lets return to the definition of  $\delta_j$ :

$$\delta_j = \frac{\partial E}{\partial n_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial n_j} = \frac{\partial E}{\partial y_j} f'(n)$$
(3.15)

Now we define  $\frac{\partial E}{\partial y_j}$  for hidden layers l :

$$\frac{\partial E}{\partial y_j} = \sum_{k=1}^{m_{l+1}} \frac{\partial E}{\partial n_k} \frac{\partial n_k}{\partial y_j} = \sum_{k=1}^{m_{l+1}} \delta_k \omega_{jk}$$
(3.16)

Thus :

$$\delta_j = \left(\sum_{k=1}^{m_{l+1}} \delta_k \omega_{jk}\right) f'(n_j) \tag{3.17}$$

From equation (3.18) we can compute local gradient layer l using local gradient of layer l+1, basing on that we can adjust the network recursively from the output layer to the input layer, this technique is used in the most common learning algorithm, Backpropagation algorithm [52, 49]

Then the adjustment of the weights in all layers is as follows :

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial\omega_{ij}} = -\eta \delta_j y_i \tag{3.18}$$

Where :

For hidden layers 
$$\delta_j = \left(\sum_{k=1}^{m_{l+1}} \delta_k \omega_{jk}\right) f'(n_j)$$
 (3.19)  
For output layers  $\delta_j = (y_i - \Omega_i) f'(n) y_i$ 

The training set  $\{\vec{x}_i, \vec{\Omega}\}\$  is then treated in two phase, the first phase is forward, consist of feeding the data to the network and compute the output without any weights adjustment,

next the backward phase starts to adjust the weights, this process is repeated for a number of epochs, hence we summarize the backpropagation algorithm as follows :

Algorithm 2 Algorithm : Backpropagation
<b>Require:</b> Training set $(\vec{x}_i, \vec{\Omega}_i)_{i=1}^k$
<b>Require:</b> Learning rate $0 < \eta < 1$
<b>Require:</b> feed-forward neural network with randomly initialized weights
Require: Number of epochs
for e in (Number of epochs) do
for all $(x_i, \Omega_i) \in (\vec{x}_i, \vec{\Omega}_i)_{i=1}^k$ do
Compute $y_i^{(i)}$ according current parameter
Compute $\delta_j$ for the output layer
Compute $\delta_j$ for Hidden layers
Update $\omega_{ij}$
EndFor
Endfor
EndFor

End

## 3.4 Regularization

Regularization is set of technique that are used in order to reduce the generalization error (overfitting) of a neural network, some of the most used regularization are : Dropout, Drop connect, Data augmentation and Stochastic depth We will focus on Dropout in the next section

## 3.4.1 Dropout

this technique refers to randomly drop neuron by removing all of its connections from the network architecture temporally during the training as discribed in figure 3.9. In other words, a neuron is present in the training phase with a probability p by consequence it is absent with a probability(1-p) [5]



Figure 3.7: applying dropout on neuron [5]



Figure 3.8: applying dropout on neural networks [5]

# 4 conclusion

In this chapter we have discussed several technique we need to construct our MLP predictive model, the way to incorporate the data set to train the machine learning model and the network adjustment using backpropagation algorithm. In the next chapter we will present the different experiments we done in this framework and discuss the obtained results as well.

# Chapter 4

# **Results and discussion**

## 1 Introduction

This chapter will be divided into two parts, the first one will be a Bayesian analysis of Judd Ofelt intensity parameters of holmium (Ho) and Erbium (Er) doped halogeno-phosphate glasses [6], we will compare and discuss the results with previous calculations namely [6, 7] In the second part we will focalize on the second technique we introduced in chapter 3, predicting Judd Ofelt parameters from glasses oxide compositions, stating by presenting the dataset we used and the machine leaning model (MLP) we established then we discuss the results.

## 2 Bayesian Inference

In this section we will begin with introducing the MCMC algorithm we used to sample from the posterior, then we will apply it to estimate the parameter of interest of glasses we mentioned in the previous section.

#### 2.1 Gibbs sampling algorithm

Since we derived the full conditionals of  $\Omega$  and  $\sigma^2$ , we can apply Gibbs sampling directly and estimate parameter of interest by sampling from the posterior thus JO parameters will be presented in term of their mean, standard deviation and probability density function Our posterior in term of full conditionals is :

$$P(\Omega, \sigma^2 \mid Y) \propto P(\Omega \mid Y, \sigma^2) \times P(\sigma^2 \mid Y, \Omega)$$
(4.1)

Where the full conditionals are :

$$P(\Omega \mid Y, \sigma^2) \sim N((X^T X)^{-1} X^T Y, \sigma^2 (X^T X)^{-1})$$
(4.2)

$$P(\sigma^2 \mid Y, \Omega) \sim IG(\frac{n}{2}, \frac{(Y - \Omega X)^T (Y - \Omega X)}{2})$$
(4.3)

Hence Gibbs sampling algorithm is as follows :

Algorithm 3 Algorithm : Gibbs samplingRequire: X = Matrix of  $(U_i^{(j)})^2$  where j = 2, 4, 6Require: Y = Vector of  $f_i^{exp}$ Require: Random initialization :  $\sigma_0^2$ Require: number of iterationfor t in (number of iteration) do $P(\Omega_t \mid Y, \sigma_t^2) \leftarrow N((X^TX)^{-1}X^TY, \sigma_t^2(X^TX)^{-1})$  $P(\sigma_{t+1}^2 \mid Y, \Omega_t) \leftarrow IG(\frac{n}{2}, \frac{(Y-\Omega_tX)^T(Y-\Omega_tX)}{2})$ end forReturn{ $\Omega_2, \Omega_4, \Omega_6$ } and  $\sigma^2$ 

We implemented the above algorithm using python with the help of its libraries such as NumPy and SciPy, however similar algorithm is integrated in PyMC3 library or RJAGS library (in R programming language)

#### 2.2 Studied samples and results

the studied glasses are halogeno phosphate glass doped with Holmium  $(Ho^{+3})$  and Erbium  $(Er^{+3})$ , systems general formula are  $(100 - x - y)NaPO_3 - xPbCl_2 - yBaCl_2 : LnF_3$  with (Ln = Ho or Er) and  $NaPO_3 - ZnF_2 - SrF_2 : HoF_3$ 

**2.2.1** System  $(100 - x - y)NaPO_3 - xPbCl_2 - yBaCl_2 : LnF_3$ 

the doping consecration of rare earth was set at 1%, glasses composition, densities and refractive index are summarized in the tables below

Name of	Glasses compositions (mol%)					
Series	NaPo <sub>3</sub>	$PbCl_2$	$BaCl_2$	$HoF_3$	$ErF_3$	
NPE1	90	9	0	0	1	
NPBE1	80	9	10	0	1	
NPH1	90	9	0	1	0	
NPBH1	80	9	10	1	0	

Table 4.1: Compositions of glass samples [6]

Camarala	Density	Refractive
Sample	$(g/cm^3)$	index
NPE1	2.901	1.525
NPBE1	3.065	1.528
NPH1	2.894	1.521
NPBH1	2.981	1.522

Table 4.2: Physical properties of glass samples [6]





Figure 4.1: Absorption Spectra of  $Er^{3+}$  doped NPE1 and NPBE1 glasses [6]

Figure 4.2: Absorption Spectra of  $Ho^{3+}$  doped NPH1 and NPBH1 glasses [6]

- Holmium cross section : For glasses doped with Holmium, the absorption spectra are presented in the figures 4.2 .There can be ten absorption bands centered at 1957, 1155, 642, 538, 486, 472, 448, 418, 386 and 360nm corresponding to the optical transitions of the  $Ho^{3+}$  ion from its  ${}^{5}I_{8}$  state to the different excited states:  ${}^{5}I_{7}$ ,  ${}^{5}I_{6}$ ,  ${}^{5}F_{5}$ ,  $5_{4}^{F}$ ,  ${}^{5}F_{3}$ ,  $({}^{5}F_{2}$ ,  ${}^{3}K_{8})$ ,  ${}^{5}G_{6}$ ,  ${}^{5}G_{5}$ ,  ${}^{5}G_{4}$  and  ${}^{3}H_{6}$  respectively. These are the ten bands that are used for the calculation of Judd-Oflet parameters
- Erbium cross section : The absorption spectra of lenses doped with erbium are presented in the Figure (4.1). From these spectra, one can count ten bands of absorptions centered at 1533, 974, 800, 651, 544, 520, 488, 451, 406 and 378nm which are attributed to optical transitions of the Er3+ ion from its  ${}^{4}I_{15/2}$  state to

the different excited state  ${}^{4}I_{13/2}$ ,  ${}^{4}I_{11/2}$ ,  ${}^{4}F_{9/2}$ ,  ${}^{4}S_{3/2}$ ,  ${}^{3}H_{11/2}$ ,  ${}^{4}F_{7/2}$ ,  ${}^{4}F_{5/2}$ ,  ${}^{2}H_{9/2}$ and  ${}^{4}G_{11/2}$ , These are the ten bands that are used for the calculation of Judd-Oflet parameters.

		Oscillator						
${}^{5}I_{8} \rightarrow$	$\lambda(\text{nm})$	strength $f_{exp}(\times 10^{-6})$						
		NPH1	NPBH1					
${}^{5}I_{7}$	1957	1,454	1.040					
${}^{5}I_{6}$	1155	0,765	0.593					
${}^{5}F_{5}$	642	2,364	1.571					
$5_{4}^{F}$	538	3,371	1.366					
${}^{5}F_{3}$	486	1,203	0.716					
${}^{5}F_{2}, {}^{3}K_{8}$	472	0,681	0.323					
${}^{5}G_{6}$	448	16,752	9.611					
${}^{5}G_{5}$	418	2,937	1.525					
${}^{5}G_{4}$	386	1,586	0.589					
$^{3}H_{6}$	360	5,234	2.281					

Table 4.3: Measured oscillator strength for  $Ho^{3+}$  ions in NPH1 and NPBH1 glasses [6]

		Oscillator						
$^4I_{15/2} \rightarrow$	$\lambda(\text{nm})$	$  \text{strength} f_{exp}(\times 10^{-6})$						
		NPE1	NPBE1					
$^{4}I_{13/2}$	1533	1,607	1,989					
${}^{4}I_{11/2}$	974	0,401	0,400					
$^{4}I_{9/2}$	800	0,159	0,167					
$^{4}F_{9/2}$	651	1,583	1,563					
${}^{4}S_{3/2}$	544	0,246	0,164					
$^{3}H_{11/2}$	520	6,473	$6,\!654$					
${}^{4}F_{7/2}$	488	1,385	1,195					
${}^{4}F_{5/2}$	451	0,233	0,272					
$^{2}H_{9/2}$	406	0,353	0,496					
$^{4}G_{11/2}$	378	11,486	11,072					

Table 4.4: Measured oscillator strength for  $Er^{3+}$  ions in NPE1 and NPBE1 glasse[6]

The tensors  $U_i^j$  are tabulated by carnal et al [53, 54], thus gibbs sampling algorithm was executed to obtain the following results.



Figure 4.5: NPE1

Figure 4.6: NPBE1

$O(\times 10^{-20})$	NPH1		NPBH	1	NPE1		NPBE1		
52 (×10)	BI	LS	BI	LS	BI	LS	BI	LS	
$\Omega_2$	$4.283 \pm 0.19$	4,280	$2.450 {\pm} 0.29$	2.442	$2.033{\pm}0.18$	2.031	$2.018 {\pm} 0.13$	2.017	
$\Omega_4$	$2.642 \pm 0.22$	2,654	$1.390{\pm}0.44$	1.407	$0.525 {\pm} 0.30$	0.525	$0.446 {\pm} 0.22$	0.449	
$\Omega_6$	$1.742 {\pm} 0.15$	1,746	$1.326 {\pm} 0.34$	1.312	$0.491 {\pm} 0.13$	0.489	$0.530{\pm}0.10$	0.531	

Table 4.5: Judd Ofelt parameters using least square method [7] and Bayesian inference method

#### **2.2.2** System $NaPO_3 - ZnF_2 - SrF_2 : LnF_3$ (Ln = Ho)

Table (4.4) gives molar composition for the System NaPO3 – ZnF2 – SrF2: LnF3 (Ln = Ho):

Composition (mol%)	NPSZ0	NPSZH0.5	NPSZH1	NPSZH1.5	NPSZH2
NaPO <sub>3</sub>	80	79.5	79	78.5	79.5
$SrF_2$	10	10	10	10	10
$ZnF_2$	10	10	10	10	10
$HoF_3$	-	0.5	1	1.5	2

Table 4.6: Compositions of glass samples [7]

$5I_{-} \rightarrow$	Osc	Oscillator strength $f_{exp}(\times 10^{-6})$										
18 -7	NPZSH0.5	NPZSH1	NPZSH1.5	NPZSH2								
${}^{5}I_{7}$	1,1608	$1,\!1951$	1,2077	1,2446								
${}^{5}I_{6}$	0,5794	$0,\!5946$	$0,\!5845$	$0,\!6287$								
${}^{5}F_{5}$	2,3700	2,3739	2,3503	2,5009								
${}^{5}F_{4}$	3,2740	3,3668	$3,\!2975$	$3,\!4516$								
${}^{5}F_{3}$	0,6234	0,8187	0,7174	0,8406								
${}^{5}F_{2}, {}^{3}K_{8}$	0,5343	0,5251	0,5275	0,5338								
$^{5}G_{6}$	11,9114	$12,\!4917$	$12,\!2562$	11,7716								
${}^{5}G_{5}$	2,2525	2,3285	2,2879	2,1654								
${}^{5}\overline{G}_{4}$	$0,\!6\overline{056}$	$0,3\overline{5}33$	$0,\!6\overline{052}$	$0,\!4\overline{850}$								
$^{3}H_{6}$	3,5322	3,4714	3,7526	3,7152								

Oscillator forces and Judd-Ofelt parameters are presented in tables (4.6) and (4.7) respectively.

Table 4.7: Oscillator strength of  $ho^{3+}$  absorption bands of NPZSHn (n=0.5, 1, 1.5, 2)[7]

The result after applying Gibbs sampling is as follows :



Figure 4.7: NPSZH0.5



Figure 4.8: NPSZH1



Figure 4.9: NPSZH1.5



Figure 4.10: NPSZH2

$\Omega~(\times 10^{-20}~)$	N	PZSH0.5	N	PZSH1	N	PZSH1.5	NPZSH2		
	LS	BI	LS	BI	LS	BI	LS	BI	
$\Omega_2$	2.243	$2.828 {\pm} 0.21$	3.035	$2.972 {\pm} 0.24$	2.983	$2.946{\pm}0.21$	2.85	$2.802 \pm 0.20$	
$\Omega_4$	3.035	$2.188 {\pm} 0.31$	2.262	$2.238 {\pm} 0.34$	2.248	$2.209 {\pm} 0.28$	2.164	$2.129 \pm 0.28$	
$\Omega_6$	1.65	$1.621 {\pm} 0.23$	1.721	$1.695 {\pm} 0.24$	1.675	$1.645 {\pm} 0.21$	1.832	$1.799 {\pm} 0.21$	

Table 4.8: Calculated Judd Ofelt parameters using least square method [7] and Bayesian inference method

#### 2.3 Discussion

The Bayesian inference framework allows us to estimate quantities of interest ( $\Omega$  parameters) from their probability densities, the fit of the 10 observed transition band of  $Ho^{+3}$  and  $Er^{+3}$  and the calculation of JO parameters by sampling from the posterior ( $P(\Omega, \sigma^2 | Y)$ ) exhibits a good accuracy with least square method with a reasonable marge of error, however when the transition bands are small the contribution of the prior may affect the accuracy of the obtained JO parameters thus in such cases we must be careful to choose the appropriate prior.

## 3 Judd Ofelt parameters prediction

In this section we will begin by presenting the data-set and data prepossessing we proceeded after that we are going to present the MLP model we established to predict the JO parameters, finally we will discuss the result we obtained in our experiment

#### 3.1 Dataset

The data consist of  $135 \times 28$  matrix, 135 observation and the first 25 features are the glasses oxide compositions, the rest are the target parameters (W2, W4, W6)

	TeO2	ZnO	Na2O	Er2O3	TiO2	BaO	Na2CO3	GeO2	B2O3	MgO	 PbO	PbCI2	PbBr2	PbF2	CaO	P2O5	W2	W4	<b>W</b> 6
<b>S1</b>	80.721157	11.931813	4.543149	2.803881	0.000000	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0	0.0	0.0	2.14	2.77	1.42
<b>S2</b>	80.651304	11.938790	4.545806	2.805520	0.058579	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0	0.0	0.0	3.08	2.15	1.88
<b>\$</b> 3	80.581370	11.945776	4.548465	2.807162	0.117227	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0	0.0	0.0	3.64	1.27	2.17
<b>\$</b> 4	80.511354	11.952770	4.551128	2.808805	0.175943	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0	0.0	0.0	3.72	1.30	2.22
<b>\$</b> 5	80.441255	11.959771	4.553794	2.810451	0.234728	0.0	0.0	0.0	0.0	0.0	 0.0	0.0	0.0	0.0	0.0	0.0	3.49	1.90	2.09

Figure 4.11: Dataset head Of bulk oxide composition percentage of  $Er^{+3}$  doped glasses and the target JO parameters

The dataset summary is given in the below table in term of count, mean, standard deviation (std), minimum, 25% quartile, 50% quartile (median), 75% quartile and maximum.

	$\operatorname{count}$	mean	$\operatorname{std}$	$\min$	25%	50%	75%	max
TeO2	135.0	64.847922	21.065089	6.934969	50.060732	71.115493	80.616337	97.257630
ZnO	135.0	4.174459	6.595924	0.000000	0.000000	0.000000	8.730498	49.512032
Na2O	135.0	1.963268	3.574073	0.000000	0.000000	0.000000	4.456195	26.931667
Er2O3	135.0	3.336365	3.710409	0.000000	1.344193	2.457325	3.377692	21.030062
TiO2	135.0	0.025896	0.231621	0.000000	0.000000	0.000000	0.000000	2.667932
BaO	135.0	1.466929	5.954325	0.000000	0.000000	0.000000	0.000000	33.463789
Na2CO3	135.0	0.083639	0.722122	0.000000	0.000000	0.000000	0.000000	7.526298
GeO2	135.0	1.165608	5.559750	0.000000	0.000000	0.000000	0.000000	43.531003
B2O3	135.0	7.591094	13.292711	0.000000	0.000000	0.000000	16.516620	60.207354
MgO	135.0	0.843718	1.948374	0.000000	0.000000	0.000000	0.000000	7.459648
K2O	135.0	1.386261	4.039041	0.000000	0.000000	0.000000	0.000000	17.436695
$\operatorname{SrO}$	135.0	0.305779	1.423135	0.000000	0.000000	0.000000	0.000000	6.927211
WO3	135.0	3.550777	11.675218	0.000000	0.000000	0.000000	0.000000	56.782274
CaF2	135.0	0.553852	1.947162	0.000000	0.000000	0.000000	0.000000	10.236133
Bi2O3	135.0	2.043205	8.801733	0.000000	0.000000	0.000000	0.000000	62.304563
CdO	135.0	0.112292	0.793789	0.000000	0.000000	0.000000	0.000000	7.573747
La2O3	135.0	1.085141	4.089506	0.000000	0.000000	0.000000	0.000000	21.560515
Li2O	135.0	0.553126	3.976812	0.000000	0.000000	0.000000	0.000000	45.225111
Nb2O5	135.0	0.950098	3.050474	0.000000	0.000000	0.000000	0.000000	16.780489
PbO	135.0	0.948978	5.074445	0.000000	0.000000	0.000000	0.000000	43.015395
PbCl2	135.0	0.499573	3.036708	0.000000	0.000000	0.000000	0.000000	29.724166
PbBr2	135.0	0.845390	6.161590	0.000000	0.000000	0.000000	0.000000	59.582157
PbF2	135.0	0.201201	2.337748	0.000000	0.000000	0.000000	0.000000	27.162173
CaO	135.0	0.500664	2.615281	0.000000	0.000000	0.000000	0.000000	19.044058
P2O5	135.0	0.964765	5.554707	0.000000	0.000000	0.000000	0.000000	36.235489
W2	135.0	5.506743	2.207121	0.900000	3.890000	5.620000	6.575000	11.990000
W4	135.0	1.760747	0.742383	0.171000	1.363500	1.610000	1.970500	5.018000
W6	135.0	1.360110	0.652335	0.020000	0.900100	1.200000	1.770000	3.540000

Table 4.9: Dataset summery



Figure 4.12: dataset heat map

## 3.2 Experiment settings

• Data Pre-processing : We scaled the dataset using the Min-Max scaling to make the features in range [-1, 1], lower the standard deviation and to suppress the effect of outliers, the general formula of Min-Max scaling in range [a, b] is :

$$x' = a + \frac{(x - min(x))(b - a)}{max(x) - min(x)}$$
(4.4)

Where x' is the scaled value of the actual value x.

- Number of test: is the number of times that we repeat the model training and evaluation to ensure that the acquired results are not exclusive to one experiment only, thus we repeated the experiment for 1000 times, each time 80% of dataset is being selected randomly as a training set
- **Batch size :** We used the whole dataset, because it is small and will not overflow the memory.

## 3.3 MLP structure

We give the below table to sum up the different hyper-parameters we used to construct our MLP predictive model

Model type	Multilayer perceptron (MLP)
Analysis type	Regression
Number of layers	5 (3 hidden layers)
	H1: 110
Hidden layer neurons	H2: 100
	H3:40
	H1 : Sigmoid
Layers activation	H2 : Sigmoid
function	H3 : ReLu
	Output layer : Linear
Loss function	Mean squared error (MSE)
Learning algorithm	Backpropagation
	H1 : No dropout
Regulariztion	H2 : Dropout rate $(0.25)$
	H3 : Dropout rate $(0.25)$
Optimizor	Adaptive Moment Estimation
Optimizer	(Adam)
Learning rate	0.001
Number of epochs	1050

Table 4.10: Summary of the options and parameters selected for our MLP model

The implementation of the model has been done using Pytorch, an open source machine learning framework with python front end, extended with other python libraries such as Scikit learn and NumPy.

As the table shows the number of neuron decreases as the number of layers increases to get the best representation of input data and the output has only 3 dimensions, dropout is used to prevent overfitting the model because the model contains fairly big number of neurons and layers in comparison to the small dataset (135 observation), Adam was used as gradient decent optimizer to speed up the training process, however this parameters are adjusted based on experimentation and observation with model and are not 100% explainable.

## 3.4 Result

As we pointed in experimental settings section, we repeated the experiment 1000 time with random selection of the training set which allocate 80% and the rest allocated for testing set. We calculated the root mean squared error to get a better observation on how much the model predictions are shifting from the target values, the RMSE for testing is given by :

$$RMSE = \sqrt{\frac{1}{14} \sum_{i=1}^{14} (\Omega - \Omega')^2}$$
(4.5)

Where  $\Omega'$  are the predicted JO parameters, for 1000 iteration the average RMSE is then :

$$\widehat{RMSE} = \frac{1}{1000} \sum_{i=1}^{1000} RMSE$$
(4.6)

The result corresponding training and testing RMSE, standard deviation, max and min RMSE values are recapitulated in the next table.

Paramotor	Tra	ining (10	$^{-20})$	Tes	Testing $(10^{-20})$			
	$\Omega_2$ $\Omega_4$		$\Omega_6$	$\Omega_2$	$\Omega_4$	$\Omega_6$		
RMSE	0.7464	0.4882	0.4213	1.4843	0.5938	0.5136		
Min RMSE Value	0.5623	0.3405	0.2913	0.7648	0.2492	0.2434		
Max RMSE Value	0.8950	0.6428	0.6044	3.6887	0.9747	0.8597		
Standard deviation	0.0537	0.0607	0.0551	0.2822	0.1376	0.0893		

Table 4.11: The estimated Root mean squared error for training and testing data from 1000 iteration of MLP regression model

#### 3.5 Discussion

the MLP predictive model we proposed to predict Judd Ofelt intensity parameters from the bulk oxide compositions, give us reasonable results from which we can infer the existence of correlation between glasses compositions and JO parameters, such relationship has no physical justification, however, during our experimentation we observe that:

- The RMSE and standard diviation of Ω<sub>2</sub> (first output) is high in comparison to the other parameters (Ω<sub>4</sub>, Ω<sub>6</sub>) RMSE .
- It is also worth noting that the model overfitts the training data fast Thus the model is significantly sensitive to the selected training and testing data.

Moreover, the selected training set notably affect the model prediction thus we can say besides small observations, the dataset contains outliers, there is undoubtedly much more to improve in this framework with more data, proper parameter distribution and include other physical factors such as refractive index, temperature and densities. Finally we point that the result we obtained aren't final.

# 4 Conclusion

In this chapter we presented the details of different tests we have done on our work, Bayesian inference and Judd Ofelt parameters prediction, with their corresponding results. We conclude that the first technique (Bayesian inference) we used is a good alternative framework in calculating JO intensity parameters, however it can be tasted more with different priors family and other MCMC simulation techniques such as Metropolis Hastings and Rejection Sampling.

The second technique we tested in our work, predicting JO parameters using Multilayer perceptron model is at first effective in detecting the correlation between quantity of interest ( $\Omega$ ) and glasses compositions, nevertheless, it can be more tested with more data and more physical properties included.

# **General Conclusion**

Judd Ofelt theory have been widely used in glasses scientific research in the past five decade, it is considered as the centerpiece of 4f shell transition spectroscopy, moreover, the theory have proven itself to be a valuable tool in the innovation of lasers, optical amplifiers, phosphors for displays and solid-state lighting, the calculation of Judd Ofelt parameters is known to be a specialized task as it require laboratory, mathematics and quantum mechanics comprehension, onece the observation bands  $f_{exp}$  are obtained, JO parameters can be determined using the well known least square methodology.

Our attempt in this thesis is to use different framework in calculation and further predicting Judd Ofelt parameters, we explained the theory behind every model and tested them in the context of our work.

The first model we used is estimating JO parameters via Bayesian inference, we fitted the observation bands  $f_{exp}$  and the matrix elements of tesors  $(U_i^{(j)})^2$  via Bayesian linear regression model, we constructed a posterior  $P(\Omega, \sigma^2 | Y)$  from which we sampled our quantity of interest ( $\Omega$ ), the estimated JO parameters of Halogeno phosphate glasses under this framework was at high accuracy with least square method thus the method was successfully tested, in fact there still tasks to appraise, therefor, we suggest :

- Using Bayesian inference method in estimating Judd ofelt parameters for other glasses systems
- Test different non-informative prior or informative priors if preliminary experimental knowledge are considered
- Use another MCMC algorithms, like Metropolis Hastings and Rejection sampling.

The second method we established is a Non-linear regression Multilayer perceptron (MLP), we tried to fit a matrix of  $(Er^{+3})$  tellurite glasses composition [43] with their corresponding JO parameters, we succeed in detecting the correlation with reasonable interval of error, yet, the first parameter of interest  $(\Omega_2)$  exhibits higher interval of error, in general the method was successful, however, there still a room of improvement. Hence, we give the following perspectives:

- Include more physical and thermal properties as features to improve the accuracy of the model
- Collect more data with proper JO parameters distribution
- Test the model with different glasses systems and different dopant rare earths
- Implement a K-fold cross-validation with ideal distribution of the targets in each fold.

At the end of this thesis we conclude that the different technique we used in the framework of Judd Ofelt theory are successful and certainly can be further enhanced.

# Bibliography

- [1] Steve Barrett and S. Dhesi. The structure of rare-earth metal surfaces. 01 2001.
- [2] Brian M. Walsh. Judd-ofelt theory: principles and practices. In Baldassare Di Bartolo and Ottavio Forte, editors, Advances in Spectroscopy for Lasers and Sensing, pages 403–433, Dordrecht, 2006. Springer Netherlands.
- [3] Shang-Da Jiang, Bing-Wu Wang, and Song Gao. Advances in Lanthanide Single-Ion Magnets, volume 164. 06 2014.
- [4] Yong-Ill Lee. Application of the judd ?ofelt theory to dy3+-doped fluoroborate/sulphate glasses. *Journal of Korean Physical Society*, 59:3300-, 11 2011.
- [5] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 06 2014.
- [6] Omar Bentouila, Aiadi Kamal Eddine, F. Rehouma, and Marcel Poulain. Spectroscopic studies of rare earth-doped halogeno-phosphate glasses. *Journal of Optoelectronics and Advanced Materials*, 15:11–12, 11 2013.
- [7] Omar Bentouila. ÉTUDE DES VERRES DOPÉS PAR DES TERRES RARES POUR LES APPLICATIONS LASERS ET AMPLIFICATIONS OPTIQUES. PhD thesis, 02 2016.
- [8] B. R. Judd. Optical absorption intensities of rare-earth ions. *Phys. Rev.*, 127:750–761, Aug 1962.
- [9] G. S. Ofelt. Intensities of crystal spectra of rare-earth ions. The Journal of Chemical Physics, 37(3):511–520, 1962.

- [10] Valdeir A. Silva, José W. Silva, Paulo C. de Morais, and Noelio O. Dantas. Juddofelt parameters via bayesian inference. *Brazilian Journal of Physics*, 48(2):119–125, Apr 2018.
- [11] BENTOUILA Omar. Etude de l'effet des terres rares dans les verres application: Lasers et amplificateurs optique. Master's thesis, 2005.
- [12] Renata Reisfeld. Spectroscopy of rare earth ions. In A. Vaseashta,
   D. Dimova-Malinovska, and J. M. Marshall, editors, Nanostructured and Advanced Materials for Applications in Sensor, Optoelectronic and Photovoltaic Technology,
   pages 77–100, Dordrecht, 2005. Springer Netherlands.
- [13] W. T. Carnall and B. G. Wybourne. Electronic energy levels of the lighter actinides: U3+, np3+, pu3+, am3+, and cm3+. The Journal of Chemical Physics, 40(11):3428–3433, 1964.
- [14] Guokui Liu. Electronic Energy Level Structure, pages 1–94. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [15] J. C. Slater. A simplification of the hartree-fock method. *Phys. Rev.*, 81:385–390, Feb 1951.
- [16] Shang-Da Jiang, Bing-Wu Wang, and Song Gao. Advances in Lanthanide Single-Ion Magnets, pages 111–141. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
- [17] W. T. Carnall, G. L. Goodman, K. Rajnak, and R. S. Rana. A systematic analysis of the spectra of the lanthanides doped into single crystal laf3. *The Journal of Chemical Physics*, 90(7):3443–3457, 1989.
- [18] G. F. Imbusch and R. Kopelman. Optical spectroscopy of electronic centers in solids, pages 1–37. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.
- [19] François Auzel. Application des transferts d'énergie résonnants a l'effet laser de verres dopés avec Er3+. Annales des Télécommunications, 24(9):363–376, September 1969.
- [20] Tucto Salinas Karem Yoli. Structural, luminescence and Judd-Ofelt analysis to study the influence of post-annealing treatment on the AIN: Tb thin films prepared by radiofrequency magnetron sputtering. PhD thesis, Pontificia Universidad Catolica del Peru, 2016.

- [21] Christiane Görller-Walrand and Koen Binnemans. Chapter 167 spectral intensities of f-f transitions. volume 25 of Handbook on the Physics and Chemistry of Rare Earths, pages 101–264. Elsevier, 1998.
- [22] S. Tanabe. Optical transitions of rare earth ions for amplifiers: how the local structure works in glass. *Journal of Non-Crystalline Solids*, 259(1):1–9, 1999.
- [23] A. Kaminskii. Crystalline Lasers: Physical Processes and Operating Schemes. Laser
   & Optical Science & Technology. Taylor & Francis, 1996.
- [24] R. Viertl. PROBABILITY AND STATISTICS Volume II: Probabilistic Models and Methods Foundations of Statistics. Encyclopedia of Life Support Systems; Mathematical Sciences. EOLSS Publishers Company Limited, 2009.
- [25] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.
- [26] Allan Gut. Introductory Measure Theory, pages 1–24. Springer New York, New York, NY, 2013.
- [27] Kolmogorov and probability theory. Arbor : Ciencia, Pensamiento y Cultura, 178, 08 2004.
- [28] Jeremy Orloff and Jonathan Bloom. Bayesian updating with continuous priors. In 18.05 Introduction to Probability and Statistics. Massachusetts Institute of Technology: MIT OpenCourseWare, Spring 2014.
- [29] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.
- [30] William Feller. An Introduction to Probability Theory and Its Applications, volume 1. Wiley, January 1968.
- [31] Bruna Wundervald. Bayesian linear regression, 06 2019.
- [32] George E. P. Box. Sampling and bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society. Series A (General), 143(4):383–430, 1980.

- [33] George C. Tiao and Arnold Zellner. On the bayesian estimation of multivariate regression. Journal of the Royal Statistical Society. Series B (Methodological), 26(2):277–285, 1964.
- [34] Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. Journal of the American Statistical Association, 91(435):1343–1370, 1996.
- [35] Telba Z. Irony and Nozer D. Singpurwalla. Non-informative priors do not exist a dialogue with josé m. bernardo. *Journal of Statistical Planning and Inference*, 65(1):159–177, 1997.
- [36] Adrian F. M. Smith. Bayesian Statistics and Efficient Information Processing Constrained by Probability Models, pages 479–490. Springer Netherlands, Dordrecht, 1977.
- [37] Jeffreys' Prior Distribution. American Cancer Society, 2014.
- [38] David Firth. Bias reduction, the jeffreys prior and glim. In Ludwig Fahrmeir, Brian Francis, Robert Gilchrist, and Gerhard Tutz, editors, Advances in GLIM and Statistical Modelling, pages 91–100, New York, NY, 1992. Springer New York.
- [39] Richard M. Feldman and Ciriaco Valdez-Flores. Markov Processes, pages 181–199.
   Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [40] Markov Chain Monte Carlo Lecture Notes, booktitle = Markov Chain Monte Carlo Lecture Notes, chapter = , pages = , doi = , url = , eprint = https://www.stat.umn.edu/geyer/f05/8931/n1998.pdf, year = November 21, 2005. University of Minnesota.
- [41] Manoj Mathew. The application of markov chain monte carlo techniques in non-linear parameter estimation for chemical engineering models. 2013.
- [42] Scott M. Lynch. The Linear Regression Model, pages 165–192. Springer New York, New York, NY, 2007.
- [43] M. Konstantinidis, E.A. Lalla, G. Lopez-Reyes, U.R. Rodríguez-Mendoza, E.A. Lymer, J. Freemantle, and M.G. Daly. Statistical learning for the estimation of judd-ofelt parameters: A case study of er3+: Doped tellurite glasses. *Journal of Luminescence*, 235:118020, 2021.

- [44] Debasish Basak, Srimanta Pal, and Dipak Patranabis. Support vector regression. Neural Information Processing – Letters and Reviews, 11, 11 2007.
- [45] Mariette Awad and Rahul Khanna. Support Vector Regression, pages 67–80. Apress, Berkeley, CA, 2015.
- [46] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. Journal of Computational and Graphical Statistics, 15(2):265–286, 2006.
- [47] Christophe Croux, Peter Filzmoser, and Heinrich Fritz. Robust sparse principal component analysis. *Technometrics*, 55(2):202–214, 2013.
- [48] John Bishop. HISTORY AND PHILOSOPHY OF NEURAL NETWORKS, pages 22–96. 01 2015.
- [49] Ján Vojt. Deep neural networks and their implementation. 2016.
- [50] K.-L Du and M.N.s Swamy. Multilayer Perceptrons: Architecture and Error Backpropagation, pages 83–126. 12 2014.
- [51] Afshin Gholamy, Vladik Kreinovich, and Olga Kosheleva. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation.
- [52] S.S. Haykin. Neural Networks and Learning Machines. Number v. 10 in Neural networks and learning machines. Prentice Hall, 2009.
- [53] W. T. Carnall, P. R. Fields, and B. G. Wybourne. Spectral intensities of the trivalent lanthanides and actinides in solution. i. pr3+, nd3+, er3+, tm3+, and yb3+. The Journal of Chemical Physics, 42(11):3797–3806, 1965.
- [54] W. T. Carnall, P. R. Fields, and K. Rajnak. Electronic energy levels in the trivalent lanthanide aquo ions. i. pr3+, nd3+, pm3+, sm3+, dy3+, ho3+, er3+, and tm3+. *The Journal of Chemical Physics*, 49(10):4424–4442, 1968.