Space plasma instrument concept and analysis using simulation and machine learning techniques

by

Guangdong Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics
University of Alberta

Abstract

Much interest has been drawn toward our near-Earth space with the advent of the space-flight era. In order to better understand this highly dynamic environment, the development of measuring instruments for use on near-Earth spacecraft has become particularly important. The current inference techniques often rely on analytic formulas with many assumptions. Some of those assumptions are not well-satisfied in experimental conditions, thus leading to uncontrolled uncertainties in the inferred physical parameters. With the development of computer hardware in the 21st century, the computational power available enables us to do new science. Particle in cell (PIC) simulations can be used to simulate the satellite and instrument interactions with space environment under various space plasma and satellite conditions relevant to near-Earth orbit. The response of the instrument to the various space plasma and satellite conditions is used to construct a synthetic solution library. Following machine learning techniques, the library can be used to create multivariate regression models based on neural networks and Radial basis functions (RBF). The advantage of using a simulation approach is that it provides uncertainties in the inferred physical parameters, and it can account for more physical processes than can be accounted for in an analytical approach. The drawback is that PIC simulations are time-consuming, whereas inferences made with analytic formulas can be much faster. The multivariate regression approach combines all those advantages: it provides uncertainties, it accounts for realistic physical processes, and it can make inferences very efficiently. Two selected instruments are studied using the simulation-based regression approach. RBF-one of the regression approach is also modified to be more efficient.

Preface

The third chapter of this thesis is based on a research article accepted for publication in Computer Physics Communication named "Inference of m-NLP data using Radial Basis Function regression with center-evolving algorithm", by Guangdong Liu and Dr. Richard Marchand. Finding optimal centers can be very time-consuming for Radial Basis Function (RBF) regression. This work describes an original algorithm that Dr. Marchand and I co-developed during my Ph.D. study, which is very efficient in finding optimal RBF centers. Dr. Marchand and I wrote two separate programs for this algorithm. I maintain my version of the program, and it accompanies the paper.

The fourth chapter is based on a research article published in the Journal of Geophysical Research: Space Physics (JGR), by Guangdong Liu and Dr. Richard Marchand. The paper's title is "Kinetic simulation of segmented plasma flowmeter response in the ionospheric plasma". This work describes an original concept of a space plasma flowmeter that can be mounted on a low Earth orbit satellite to measure the transverse plasma flow velocity.

The fifth chapter is based on another research article submitted to JGR in 2022 entitled " m-NLP inference models using simulation and regression techniques". This work is done in collaboration with colleagues in Norway, namely, Sigvald Marholm and Lasse Clausen from the University of Olso, and Anders Eklund from the Institute for Energy Technology. I was fully responsible for doing the simulations, training and validating the regression models, applying the models to simulation and in-situ data, writing the original manuscript, and editing it. Dr. Richard Marchand supervised the

project, and was involved in concept formation, manuscript composition and editing. Sigvald Marholm and Anders Eklund assisted in conceptualization and manuscript editing. Lasse Clausen contributed to in-situ data collection and providing feedback in writing of the manuscript. This work presents a new approach to interpreting multi-needle Langmuir probe (m-NLP) measurements over the currently used state-of-the-art analytic approaches. Our approach, based on results from 3-D particle in cell simulations accounts for more physical processes than possible in theories, will improve the accuracy of the current m-NLP measurements and provide qualitative uncertainties.

Acknowledgments

This pandemic that started during my second year of study really changed many things in life. I want to first thank my supervisor Dr. Richard Marchand. He works very hard, and since we meet twice a week, seeing him working so hard really motivates me to work hard as well. He also helped me a lot in editing papers and refining research strategies. I could not have completed my thesis without his help. I want to thank my supervisory committee members, Dr. Fernandez and Dr. Meldrum for their support. I also want to thank my good friend Akinola Olowookere. It is good to have someone to discuss problems and help each other. I want to thank my co-authors in Norway, Sigvald Marholm, Anders Eklund, and Lasse Clausen for their valuable discussions and input in the manuscript entitled "m-NLP inference models using simulation and regression techniques". My deepest gratitude also goes to my father and my mother for encouraging and supporting me through the years. The simulation results were obtained using Compute Canada computing facilities. Last but very importantly, I want to thank China Scholarship Council for its financial support.

Table of Contents

List of Tables

List of Figures

Abbreviations & Acronyms

BNLF Barjatya nonlinear fit.

CERBF center-evolving Radial Basis function.

IRI International Reference Ionosphere.

JLF Jacobsen linear fit.

MAE Maximum absolute error.

m-NLP Multi-needle Langmuir probe.

MPI Message Passing Interface.

MRE Maximum relative error.

OML Orbital Motion Limited.

PIC Particle in cell.

RBF Radial Basis function.

RMS Root mean square error.

RMSr Root mean square relative error.

SF meter Segmented flow meter.

Chapter 1

Introduction

1.1 Motivation

The 20st century is an era of the digital revolution. According to Moore's Law, the number of transistors per silicon chip doubles every year. We have never before had this much computational power available. This enables us to apply advance techniques to solve real life problems in all field of science, including simulation of complex physical processes in a virtual environment. One objective of this thesis is to take advantage of the computational power available today to do new science.

With the advent of the space-flight era, much interest has been drawn toward our near-Earth space. The highly dynamic environment influenced by space weather and electric and magnetic fields around Earth makes it particularly difficult to study. One reason is that instruments such as Langmuir probes rely on theoretical models involving many simplifications and assumptions. When these assumptions are not fully satisfied, the reliability of such models becomes questionable. It is also difficult to calibrate an instrument in space. In order to do a calibration, there must be another instrument to measure the same quantity, which is difficult in space considering the cost of a spacecraft, and the calibration instrument can have uncertainties as well. Thus it is especially cost-efficient to calibrate the instruments in labs or, more conveniently using simulation techniques prior to launch.

1.2 Thesis objectives and outline

In this thesis, I studied selected near-Earth satellite instruments using kinetic simulations and assess their measurement skills based on simulation results. More specifically, the instruments considered are the multi-needle Langmuir probe (m-NLP) system onboard the NorSat-1 satellite and a new type of plasma flowmeter. The specific objectives are to:

1. Construct and assess a new regression-based model for m-NLP instrument based on simulation results.

2. Compare the inferences made with various state-of-the-art m-NLP models with those made with regression models based on simulation and in-situ data.

3. Propose a new concept of a plasma flowmeter, and determine its measurement uncertainties based on simulation techniques.

4. Improve the efficiency of the Radial Basis Function algorithm to be able to construct models using large data sets.

The remainder of this thesis is structured into six main chapters. In Chapter 1, the general theme and background of the thesis are discussed. In chapter 2, the general methodology is described. An improvement to the multivariate Radial Basis Function (RBF) is presented in Chapter 3, where it is demonstrated that the new technique can be significantly more efficient than the previously used straightforward approach, at no cost in accuracy. A review of space plasma flow meter designs is presented in Chapter 4, along with a proposed new design for a simple and robust flow meter. Chapter 5 presents a new approach to infer plasma and satellite parameters using m-NLP, based on simulations and regression techniques. The model is first assessed on the basis of synthetic data, and then applied to in-situ measurements

reported from the Norwegian NorSat-1 satellite. The currently used inference model for needle Langmuir probe is often based on the Orbital Motion Limited (OML) theory derived nearly a century ago based on several simplifications. These include the assumption that the needle probe is much larger than the Debye length, that the plasma is stationary; conditions which are generally not well satisfied for actual needle probes used on satellites in low Earth orbit. A promising alternative to the OML theory is to construct inference models based on computer simulations in which actual dimensions of probes are taken into account, in more realistic space plasma environment conditions. In this work, currents collected by an m-NLP instrument are simulated under a variety of plasma and satellite conditions, which are then used to build models to infer plasma density and satellite floating potentials. The models are applied to in-situ data. While pure OML-based models fail to make reasonable inferences from in-situ data, simulated-based model inferences are found to be much more consistent with satellite observations. Chapter 6 includes some concluding remarks and discussions of possible future work.

1.3 Background

Understanding our near-Earth space environment has been of strategic importance due to applications such as communications, GPS positioning, monitoring of solid Earth activities, and satellite maintenance. This highly dynamic environment is bombarded by energetic particles and solar UV, causing some of the valence electrons to be liberated from the air molecules once they gain enough energy. This process is called ionization, and an ionized gas is called plasma which is the main ingredient of our upper atmosphere. Ionization of upper atmosphere atoms and molecules happens mainly in the D, E, and F regions at altitudes of approximately 70 to 400 km. Once ionized and energized, particles expand to higher altitudes beyond the ionosphere, up to approximately 2000 km in the plasmasphere. This layer of plasma within the magnetosphere co-rotates with the Earth.

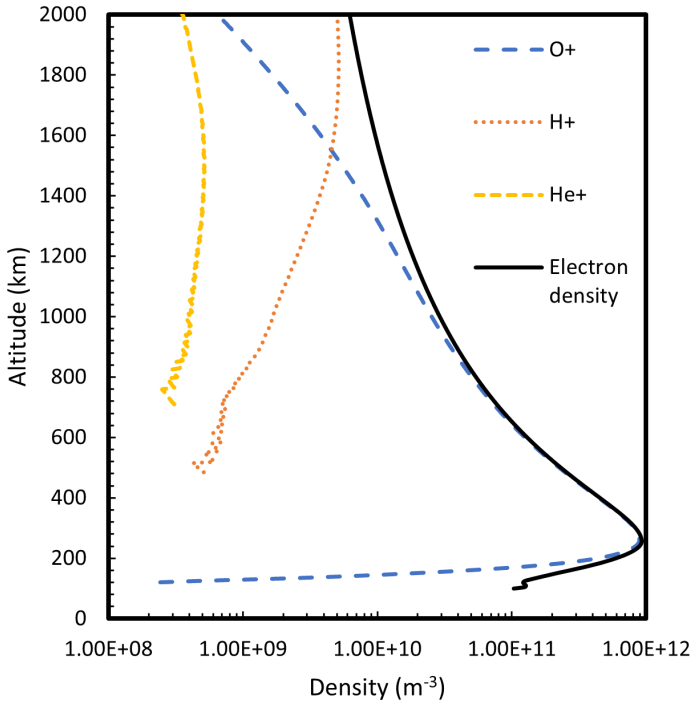

Figure 1.1: Illustration of electron and selected ion density profiles in the upper atmosphere and ionosphere.

Two key parameters in plasma physics are the electron Debye length:

$$\lambda_{De} = \left(\frac{\varepsilon_0 k T_e}{n_e e^2} \right)^{\frac{1}{2}}, \tag{1.1}$$

and plasma frequency:

$$\omega_{pe} = \left(\frac{n_e e^2}{m_e \varepsilon_0} \right)^{\frac{1}{2}}. \tag{1.2}$$

In these equations, ε_0 is the free space permittivity, n_e is the plasma density, e is the elementary charge, k is the Boltzmann constant, T_e is the electron temperature, and m_e is the mass of an electron. The Debye length is a measure of how far a perturbation of electric field can extend in a plasma. For example, if a charged object is placed in plasma, opposite charges will be attracted near it and shielding it so that other charged particles in a few Debye lengths will not be affected by the presence of this object. Lighter electrons in plasma are more mobile than heavier ions. If plasma is perturbed by an external force, electrons will be accelerated back and forth around their equilibrium positions resulting in collective oscillations. The frequency

of this oscillation, called the plasma frequency, depends on the electron charge, mass, and plasma density. There are three requirements for an ionized gas to be called plasma [1]. First, the dimension of the system must be larger than the Debye length. Secondly, there must be many charged particles inside a sphere of one Debye length radius. Lastly, the average time between electron-neural collisions multiplied by the plasma frequency must be much larger than one. In other words, electrons must be mostly unaffected by collisions.

1.3.1 Multi-needle Langmuir probe instruments

Langmuir probes have been used in labs and on spacecraft to infer plasma parameters such as density and temperature. A Langmuir probe is a conductor, which can be spherical, cylindrical, or planar. Once immersed in plasma with an applied voltage, it can be used to measure plasma current. A commonly used approach to measure plasma currents is to sweep the bias voltage on the probe, and measure the current as a function of bias voltage. The resulting current as a function of bias voltage is called the probe characteristic. An example characteristic is shown in Fig. 1.2. The ion saturation region is where the probe is negatively biased and mainly collects ions. In the electron or ion retardation region, the probe bias voltage is close to zero, thus collecting both electrons and ions. In the electron saturation region, the probe is positively biased and collects mainly electrons. The period used in sweep mode mainly depends on two factors: sensitivity of the instrument, and conditions in the surrounding plasma. In particular, short sweep periods are only possible in plasma with a sufficiently high density for currents to be measured with an acceptable signal-to-noise ratio. The sweep period varies based on the conditions, typically on the order of hundreds of milliseconds to seconds for near-Earth satellites [2, 3]. Considering the orbital speed of a spacecraft in a near-Earth orbit is on the order of 7500 m/s, the spacecraft can travel tens to thousands of meters during one sweep which limits the spatial and temporal resolutions of the instrument. In order to

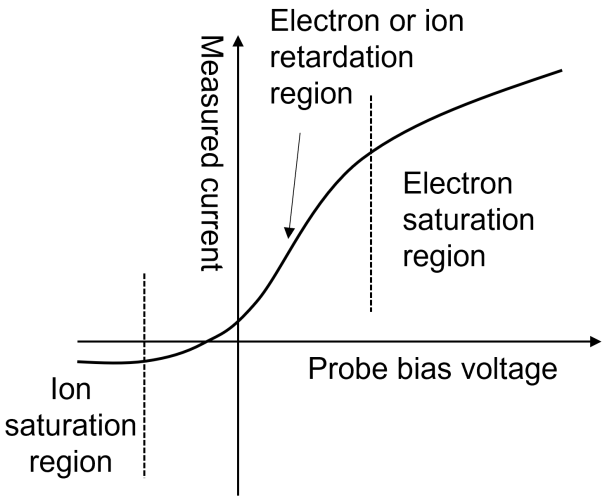

Figure 1.2: Illustration of an IV characteristic curve for a Langmuir probe.

increase the sampling frequency, in 2010, Jacobsen suggested using a multi-needle Langmuir probe (m-NLP) instrument that can be used to sample plasma at different probe voltages simultaneously by having multiple fixed bias probes. Compared to a sweep mode Langmuir probe, m-NLP instruments can have a sampling rate over kHz range which is particularly important for the study of irregularities caused by various plasma instability processes at high and low latitudes [3, 4]. The inhomogeneous magnetic-field aligned density irregularities scale from meters to tens of kilometers may interfere with the radio wave frequency band used in radio communication [4, 5]. Thus it is crucial to be able to study them and forecast the events that can cause disturbances in radio communication and global navigation satellite signals [4]. The Norwegian satellite NorSat-1 launched on July 14th, 2017 is equipped with an m-NLP instrument consisting of four needle probes biased at 6, 8, 9, and 10 volts with respect to the spacecraft [6]. The instrument can sample the surrounding plasma at a rate up to 1 kHz. The high sampling rate of the NorSat-1 m-NLP instrument allows a spatial resolution of a few tens of meters, which is sufficient to study such phenomena and help to understand instability driving mechanisms. Despite these advantages, the interpretation of m-NLP measurements in terms of plasma parameters has proven to be more challenging than initially anticipated based on simple analytic models. Thus,

the primary objectives of this thesis are to find alternatives to analytic formulas which provide higher accuracy and uncertainties in the inferred physical quantities.

1.3.2 Multi-needle Langmuir probe theory

There are different theories for cylindrical Langmuir probes, such as the Orbital Motion-Limited (OML) theory, the Allen-Boyd-Reynolds (ABR) theory, and the Bernstein-Rabinowitz-Laframboise (BRL) theory. For example, the ABR theory relies on the solution of Poisson's equation to determine the electric potential in the vicinity of a probe. The electrons are assumed to be Maximilian, whereas ions are assumed to be cold. The ions initially at $r = \infty$ are drawn radially into the probe without orbital motion [7], and Poisson's equation is solved to calculate the probe potential at the probe radius $r = R$ for a given normalized probe current. Conversely, the BRL theory accounts for both sheath formation and orbital motion [7]. In this thesis, I focus mainly on the OML theory which was derived in the 1920s by Langmuir and co-workers [8], and it is still used routinely for lab plasma and spacecraft data analysis. In this theory, a collisionless plasma and purely radial force are assumed, thus conservation of energy and angular momentum can be used to solve for the collision impact parameter. The impact parameter b is the perpendicular distance between the extended initial path of the incoming particle to the center of the cylinder as shown in Fig. 1.3. Assuming the cylinder is positively biased with respect to the background plasma and it is in the electron saturation region, the equation for conservation of energy for an incoming particle is given by:

$$\frac{1}{2}mv_0^2 = \frac{1}{2}mv_t^2 - eV, \tag{1.3}$$

and the equation for conservation of angular momentum by,

$$mv_0 b = mv_t r, \tag{1.4}$$

where r is the radius of the cylinder, v_0 is the initial speed of the particle, v_t is the particle speed at a distance r from the probe center, m is the mass of the particle, and

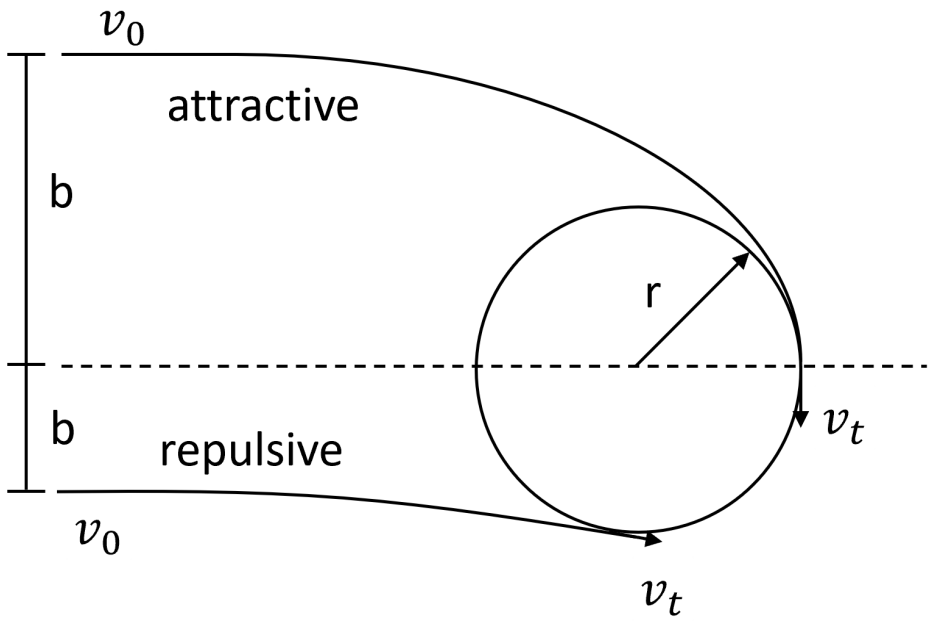

Figure 1.3: Illustration of attractive and repulsive charged particles moving towards a cylindrical probe that is perpendicular to the page.

V is the probe voltage with respect to the background plasma. The two equations can be used to solve for the impact parameter b:

$$b = r\sqrt{1 + \frac{2eV}{mv_0^2}}. \tag{1.5}$$

In order to calculate the amount of current collected by the cylinder, a Maxwellian velocity distribution is assumed for the background plasma. The two-dimensional Maxwellian distribution function has the following form:

$$f(\vec{v}) = n\frac{m}{2\pi kT} exp\left(-\frac{m \mid \vec{v} - \vec{v}_d \mid^2}{2kT}\right), \tag{1.6}$$

where n is the density, m is the mass of the particle, and $\vec{v}_d$ is the drift velocity. The particle flux density on a surface is given by:

$$\Gamma = \int \mathrm{d}^3 v f(\vec{v}) \vec{v} \cdot \hat{n}. \tag{1.7}$$

Ignoring drift, the current collected on the cylinder can be calculated using the following formula:

$$I = e \cdot \Gamma \cdot A = e2\pi l \int_0^\infty v \mathrm{d}v f(\vec{v}) 2bv, \tag{1.8}$$

where $A = 2\pi rl$ is the surface area of the cylinder ignoring the ends, and l is the length of the cylinder. Thus the current collected by a cylinder for an attractive species can be calculated by combining Eq. 1.5, Eq. 1.6, and Eq. 1.8:

$$I = ne2r\frac{m}{2\pi kT}2\pi l\int_0^\infty \mathrm{d}v v^2 \exp\left(-\frac{mv^2}{2kT}\right)\sqrt{1+\frac{2eV}{mv^2}}. \tag{1.9}$$

With a change of variables:

$$x = \frac{v}{\sqrt{2kT/m}}, \tag{1.10}$$

Eq 1.9 becomes:

$$I = 2nerl\frac{m}{kT}\left(\frac{2kT}{m}\right)^{\frac{3}{2}}\int_0^\infty \mathrm{d}x x^2 \exp\left(-x^2\right)\sqrt{1+\frac{eV}{kTx^2}}. \tag{1.11}$$

This can be rewritten as:

$$I = 4nerl\sqrt{\frac{2kT}{m}}\int_0^\infty \mathrm{d}x x \exp\left(-x^2\right)\sqrt{x^2+\frac{eV}{kT}}. \tag{1.12}$$

To evaluate the integral, let $a^2 = \frac{eV}{kT}$, and $u^2 = x^2 + a^2$, thus $x\mathrm{d}x = u\mathrm{d}u$, and the integral becomes:

$$\exp\left(a^2\right)\int_a^\infty u^2 \exp\left(-u^2\right)\mathrm{d}u. \tag{1.13}$$

Integrating by parts, using $f = u$, $g' = u\exp\left(-u^2\right)\mathrm{d}u$, and the fact that $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp\left(-t^2\right)\mathrm{d}t$, $\mathrm{erf}(0) = 0$, $\mathrm{erf}(\infty) = 1$ and $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$, we get:

$$\begin{aligned}
&\exp\left(a^2\right)\left(-\frac{u\exp\left(-u^2\right)}{2}\bigg|_a^\infty - \int_a^\infty -\frac{\exp\left(-u^2\right)}{2}\mathrm{d}u\right)\\
&= -\frac{u\exp\left(a^2-u^2\right)}{2}\bigg|_a^\infty + \exp\left(a^2\right)\frac{1}{2}\left(\int_0^\infty \exp\left(-u^2\right)\mathrm{d}u - \int_0^a \exp\left(-u^2\right)\mathrm{d}u\right)\\
&= \frac{a}{2} + \exp\left(a^2\right)\frac{\sqrt{\pi}}{4}\left(\mathrm{erf}(\infty) - \mathrm{erf(a)}\right)\\
&= \frac{a}{2} + \frac{\exp\left(a^2\right)\sqrt{\pi}\mathrm{erfc}(a)}{4}
\end{aligned} \tag{1.14}$$

This expression is found to be practically equal to $\frac{1}{2}\left(1+a^2\right)^{\frac{1}{2}}$ for $a > 1.5$ as shown in Fig 1.4, and reported in [9]. Hence, $\frac{eV}{kT} = \sqrt{a}$ needs to be larger than two for the approximation to be valid. Eq. 1.12 can then be rewritten as:

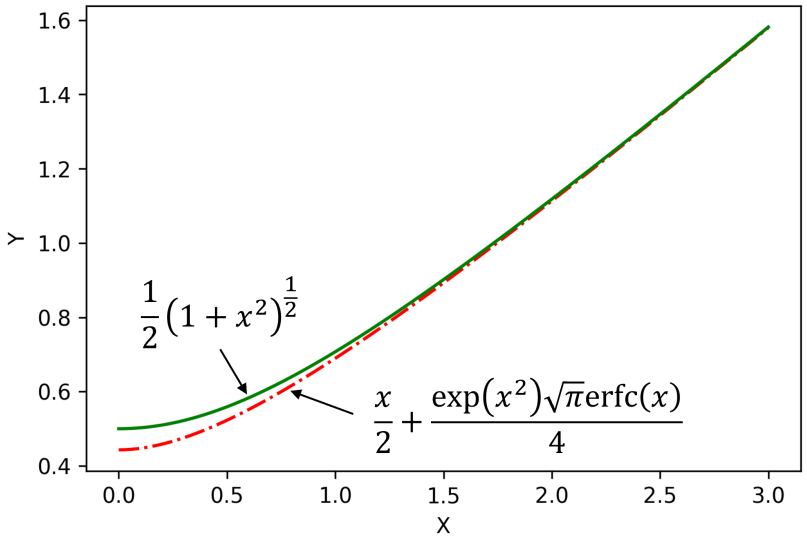

Figure 1.4: Eq. 1.14 is plotted against $\frac{1}{2}\left(1+a^2\right)^{\frac{1}{2}}$. The two expressions are practically equal for $a > 1.5$.

$$I = 4nerl\sqrt{\frac{2kT}{m}}\frac{1}{2}\left(1+\frac{eV}{kT}\right)^{\frac{1}{2}}. \tag{1.15}$$

Using the fact that $A = 2\pi rl$, this can be simplified to:

$$I = neA\sqrt{\frac{2kT}{\pi^2 m}}\left(1+\frac{eV}{kT}\right)^{\frac{1}{2}}. \tag{1.16}$$

Many studies have been conducted to investigate the performance of the above equation and try to improve it over the years since it was first introduced in the 1920s. The exponent term in this equation often written as β is only equal to $\frac{1}{2}$ if all underlying assumptions in the OML theory are satisfied, which is rarely the case in practice. Empirically, it is found that collected current as a function of bias voltage can be better approximated using a β value between 0.5 and 1. Thus replacing the $\frac{1}{2}$ with β, and the V with $V_f + V_b$ where V_f is the satellite floating potential, and V_b is the bias potential of the probe with respect to the spacecraft, a more generally used form of Eq. 1.16 is given by:

$$I = neA\sqrt{\frac{2kT}{\pi^2 m}}\left(1+\frac{e(V_f+V_b)}{kT}\right)^{\beta}. \tag{1.17}$$

For an isolated surface in plasma, the floating potential is the potential at which the surface collects net zero current. The density, temperature, floating potential and β can be obtained using these variables as fitting parameters to fit the measured current as a function of sweep bias voltage in the electron saturation region [10]. This approach is discussed in more detail in Chapter 5.

As first suggested by Jacobson, if the OML assumptions are satisfied, and $\beta = 0.5$, a set of multi-needle Langmuir probes (m-NLP) can be used to infer plasma densities independently of temperature. This is achieved by having multiple needle probes biased at different fixed bias voltages in the electron saturation region. For example, with two probes biased at V_{b1} and V_{b2}, assuming $\beta = 0.5$, the current squared collected by probe one is given by:

$$I_1^2 = \left(n_e e A \frac{2}{\sqrt{\pi}} \sqrt{\frac{kT}{2\pi m_e}} \right)^2 \left(1 + \frac{e(V_f + V_{b1})}{kT} \right). \tag{1.18}$$

The current squared collected by probe two is given by:

$$I_2^2 = \left(n_e e A \frac{2}{\sqrt{\pi}} \sqrt{\frac{kT}{2\pi m_e}} \right)^2 \left(1 + \frac{e(V_f + V_{b2})}{kT} \right). \tag{1.19}$$

Subtracting Eq. 1.19 from Eq. 1.18 to get:

$$I_1^2 - I_2^2 = \left(n e A \frac{2}{\sqrt{\pi}} \sqrt{\frac{kT}{2\pi m}} \right)^2 \frac{e}{kT} \left(V_{b1} - V_{b2} \right), \tag{1.20}$$

which can be rewritten as:

$$n = \sqrt{\frac{\pi^2 m}{2e^3 A^2}} \sqrt{\frac{I_1^2 - I_2^2}{V_{b1} - V_{b2}}}. \tag{1.21}$$

This equation can be used to calculate the density without the knowledge of the electron temperature or spacecraft potential. If there are more than two probes, the slope of the current squared vs. bias voltage can be used to calculate the plasma density. The main problem with such an approach is that the β value is rarely equal to 0.5 due to the end effect [4, 11–16]. This will be discussed in the following section and in Chapter 5.

1.3.3 Orbital Motion Limited assumptions

Several assumptions are made in the OML derivation in order to obtain an analytic formula, these include:

- The plasma is collisionless, whence the energy and angular momentum of the incoming particles are conserved.
- The plasma is non-drifting.
- Magnetic fields are negligible.
- The Debye length is much larger than the probe radius.
- The probe length is much larger than the Debye length.
- No other objects are in the vicinity of the probe.

Another condition required to simplify the OML formula is that the eV is larger than kT by at least a factor of two. Experimentally, several of these assumptions are rarely satisfied, and the theory needs to be adjusted for different experimental environments. For example, Laframboise and Rubinste extended OML theory for spherical and cylindrical probes to incorporate magnetic fields [17, 18]. When a magnetic field is present, instead of having a straight trajectory, an isolated particle will do helical motion like in $E \times B$ drift. The current collected by a probe will be influenced by the particle's gyro radius and the angle of the probe with respect to the magnetic field. A typical gyro radius of electrons ($\sim$ 20-40 mm) in ionospheric plasma is much larger than the needle probe diameter ($\sim$ 0.5 mm) considered in this study. Whence, the OML assumption of a nearly straight trajectory (oppose to a helical trajectory, see Fig. 6 in [18]), is valid; which justifies the neglect of the magnetic field in this work. For a large radius probe or a probe in a strong magnetic field, however, the magnetic field should be taken into account.

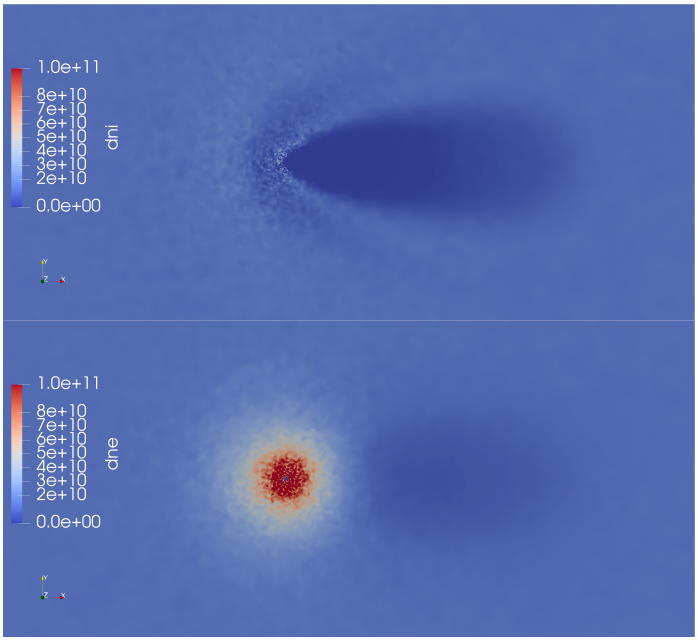

Figure 1.5: Illustration of ion (top) and electron (bottom) rarefied wake region created by a cylindrical probe that is perpendicular to the page moving to the left. Both cases are taken from the same simulation corresponding to 7500 m/s, +9 V, 1.5×10^{10} m^{-3}, 2 amu, 0.10 and 0.08 eV for orbital speed, probe voltage, density, ion effective mass, electron and ion temperatures, respectively.

At typically ionospheric temperature, the thermal speed of electrons ($\sim 1.9 \times 10^5$ m/s for 0.1 eV electron) is much higher than the orbital speed of a satellite ($\sim$ 7500 m/s) plus the ionospheric wind speed ($\leq$ 1000 m/s) in near-Earth orbits. In comparison, the thermal speed of ions ($\sim$ 4400 m/s and 1100 m/s for 0.1 eV hydrogen and oxygen, respectively) is usually lower than the orbital speed of a near-Earth satellite plus the ionospheric wind speed. Based on these estimates, one would expect that the electrons can approach the probe from all directions, however, this is not the case. When a probe is charged positively to collect electrons, ions are repelled, and form an ion rarefied wake region downstream as shown in Fig. 1.5. Electron collection from the wake region of the probe is therefore reduced due to the resulting repulsive negative space potential. Thus the electron saturation current is also influenced by an orbital speed.

Another essential assumption made in the OML theory is that the needle probe needs to be very long. Due to the end effect, the current collected on a needle probe is not uniform along the probe. Due to symmetry, a probe that is much longer than the Debye length will collect uniform current in its mid section, and the current collected in this section approaches the value predicted by the OML theory for an infinitely long probe. Near the two free ends, however, more current is collected and features characteristic peaks. If the probe is short, the peaks in the two free ends will merge into one another, and produce a single peak in the middle of the probe as shown in Fig 3 in [15]. If a guard is used, the characteristic peak will also be attenuated on the side attached to the guard. In Fig. 1.6, the probe length is 2.5 cm, and the Debye length is 2 cm; the probe length is comparable to the Debye length. Thus, we can expect to see a single peak in the middle of the probe. In order to attenuate the end effect, the needle probe needs to be long enough so that the uniform current collected in the midsection of the probe makes up the largest portion of the current collection. Different studies suggested that in order for β value to be approximately 0.5, the probe needs to have a length that is more than 30 times the Debye length as reported in [15] and also in Chapter 5. However, the majority of the needle probes used in near-Earth orbit have lengths comparable to or sometimes smaller than the Debye length. In this case, the measurements are subject to end effects, thus, the accuracy of such measurements using OML interpretation is compromised. With probes already deployed in space, we need to find alternatives to interpret their measurements. In chapter 5 of this thesis, different regression methods based on kinetic simulations of the probe's accurate dimension to interpret m-NLP data are presented.

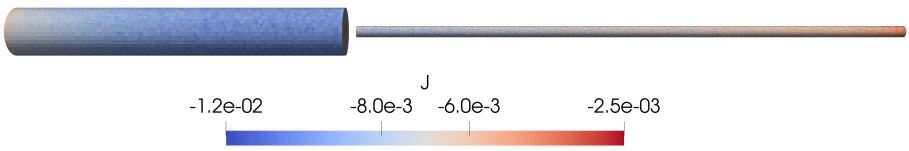

Figure 1.6: Illustration of measured current distribution on a needle probe with a guard. The simulation inputs are 7500 m/s, +9 V, 2×10^{10} m^{-3}, 8 amu, 0.15 and 0.12 eV for orbital speed, probe voltage, density, ion effective mass, electron and ion temperatures respectively.

Chapter 2

Methodology

The methodology is an essential aspect of this thesis since part of my contribution is to create a new algorithm for Radial Basis Function (RBF) multivariate regression. Although the methodology is briefly discussed in each chapter, it is beneficial to describe it in detail in a separate chapter.

We understand physics around us by creating mathematical models. For example, Newton's first law, $F = ma$, describes the motion of a moving object. With the help of a computer, we can create more complicated models. For example, we can simulate the motion of every single plasma particle inside a domain and calculate the amount of current collected on a probe. We can vary the different parameters in the simulation, for example, plasma density, electron and ion temperatures, potential of the probe with respect to the background plasma, flow speed of the plasma, ion compositions in the plasma, magnetic field, etc. Then inference models can be created for plasma parameters such as density and temperature, from probe characteristics. This is challenging since the current collected depends on many of the conditions mentioned above. However, advanced theoretical or computational models are never exact, and uncertainties in the inferences will result and need to be quantified.

The obvious advantage of simulation is that it is efficient and less time-consuming

than, for example, building a plasma chamber and modifying it for different experimental conditions. Also, a measuring instrument needs to be calibrated independently by other instruments with known confident intervals. In simulations, input physical parameters are known exactly as opposed to an assumed conditions in experiment, which needs to be measured and come with uncertainties. Furthermore, for structure stress tests, simulations are safer and will not create safety hazards in case of failure. In many cases, simulations are also fast and less expensive than doing large scale experiments in which different configurations have to be considered. For simulations, multiple trials for different configurations can be running at the same time. Of course, simulations come with limitations as well. For example, background plasma particle can have a Maxwellian or kappa velocity distribution, but these are only approximations to actual distributions encountered in space. Also in simulations, probes are generally treated as ideal conductors while neglecting alterations resulting from contamination and aging. Thus ultimately, simulations need to be validated with experiments, but by using simulations and experiments, fewer validation experiments are needed than if only experimentation were used. In contrast to a theoretical model which needs to be re-derived, simulation models can be improved by incrementally including more physical processes. Additionally, a computer can be programmed to create simulation trials and find optimal answers automatically and tirelessly.

2.1 Particle in cell simulations

The kinetic simulations in this thesis are made using particle in cell simulation code PTetra [19, 20]. The numerical approach used in PTetra is briefly discussed here, readers interested in more technical descriptions of the code can find those in the references above.

In PTetra, space is discretized using unstructured adaptive tetrahedral meshes.

Although unstructured meshes are more complex than structured meshes, they can represent objects or boundaries of irregular shapes while a structured mesh cannot. The meshes in PTetra are generated using `gmesh` [21, 22] with Delaunay connectivity. Excluding possible degeneracy in which more than four vertices are on a given sphere, the Delaunay connectivity is known to be unique and is "as compact as possible".

Due to a large number of particles in a simulation domain, simulating all particles is generally impractical. In PTetra, macro-particles, or simulation particles, are used such that each of them can represent multiple physical particles. The macro-particles have a statistical weight $\omega_\alpha = \frac{V_d n_\alpha}{N_\alpha}$, where V_d is the total volume of the simulation domain, n_α is the density of the species, and N_α is the total number of the macro-particles of this species. Physical quantities such as density and temperature, can be calculated from moments of simulated particle distributions with their respective statistical weights.

Prior to starting simulations, the range of plasma conditions relevant to the instrument's working environment must be determined. For near-Earth spacecraft, this can be done with the International Reference Ionosphere [23], to estimate plasma conditions encountered over prescribed times and locations along a satellite orbit. In order to initialize a simulation, the domain must be occupied with particles. A straightforward way to distribute particles would be to assign a random location for each particle. However, the particle distribution created using this approach would not be uniform enough, e.g. some cells might be empty. This in turn would lead to large transients at the beginning of a simulation. In order to ensure a more uniform initial distribution of particles, and minimize initial transients, PTetra uses a method such that each tetrahedron τ of volume V_τ, is assigned $\mathrm{int}\left(N_\alpha \frac{V_\tau}{V_d}\right)$ particles, where N_α is the number of total particles, and V_d is the total volume of the simulation domain. If there are still particles left after the procedure, another iteration is performed in

which single particles are added to randomly selected tetrahedra. This approach can create a much more uniformly occupied simulation domain.

Particles must also be assigned an initial velocity. PTetra can account for multiple species, each one with a given density, temperature, and drift velocity. In addition, for ions, different charges and masses can be accounted for. PTetra assumes Maxwellian velocity distributions for each plasma particle species. A random velocity can be assigned using the cumulative distribution function for the one-dimensional Maxwellian distribution function:

$$F(v) = \frac{1}{2}\left[1 + \mathrm{erf}(v - v_0)/v_{th})\right], \tag{2.1}$$

by making $F(v) = r$, where r is a random number between 0 and 1, and solve for v using a root finder. In this case, v_0 is the drifting speed, and $v_{th} = \sqrt{2kT/m}$ is the thermal speed. Note that PTetra assumes the velocity normalized probability density function is separable, that is $f(\vec{v}) = f_x(v_x)f_y(v_y)f_z(v_z)$. Thus, this procedure needs to be done three times in order to assign velocities for x, y, and z directions. Alternatively, PTetra uses a simpler method to assign a random velocity based on the central limit theorem:

$$v_\beta = v_{0\beta} + \sqrt{\frac{12kT}{Nm}} \sum_{i=1}^{N} (r_i - 0.5). \tag{2.2}$$

In this case, T is the temperature, m is the mass of the particle, k is the Boltzmann constant, r_i is a random number between 0 and 1, and N is the number of random numbers in the summation. A good approximation to a Maxwellian distribution is obtained with $N \geq 12$. In practice, $N = 96$ produces an excellent approximation.

In PTetra, particles are injected at the boundaries at each time step. The number of particles to inject is calculated by integrating the flux of a Maxwellian velocity distribution function over the boundary surface [19, 20]. The point of injection is generated randomly on the triangular surface, and a random velocity is assigned ac-

cording to a Maxwellian velocity distribution. In a time step, a particle might leave the simulation domain. Unlike a periodic boundary condition, PTetra uses an open domain, thus the particles leaving the system are lost. As a result, the number of particles for a given species can vary in time.

In order for simulations to be numerically stable, the mesh spatial resolution should be high enough to resolve the Debye length, and the time step should be small enough to resolve the plasma frequency. When a magnetic field is accounted for in a simulation, the time step should also resolve the electron gyrofrequency. Finally, in order to obtain simulation results with an acceptable signal-to-noise ratio, simulation cells (tetrahedra) should typically contain an average of 100 simulation particles or more.

Determining the electric field to update particle positions in PTetra is discussed in full in the literature [19, 20], and will be discussed briefly here. In order to calculate the electric field, the first step is to solve Poisson's equation for the electric potential. There are several contributions to a spacecraft and surrounding plasma potential. These are i) volume charge densities, ii) potential differences resulting from the motional electric field $\vec{E}_d = -\vec{v}_d \times \vec{B}$ where $\vec{v}_d$ is the plasma drift velocity in the spacecraft reference frame, iii) spacecraft charging, and iv) different bias voltages applied to spacecraft components. In order to have a well-posed problem, the potential has to be specified on every boundary, including the simulation outer boundary. Assuming a boundary sufficiently far from the spacecraft, the potential there is determined only by the motional electric field, and it is given by

$$V(\vec{r}) = \vec{v}_d \times \vec{B} \cdot \vec{r}. \tag{2.3}$$

Thus in the absence of an external magnetic field, the potential vanishes identically at the outer boundary. The boundary conditions on each spacecraft component can be calculated using the mutual capacitance matrix and the linear superposition principle

using the collected charges, the space charge density, and the outer boundary conditions [20]. To calculate the space charge density ρ, a linear interpolation function $\psi_j(\vec{r})$ must be defined at every vertex j:

$$\psi_j(\vec{r}) = \begin{cases} 1 & \text{where } \vec{r} = \vec{r_j} \\ 0 & \text{at every other vertex,} \end{cases} \tag{2.4}$$

for all tetrahedra in the simulation domain. The function ψ_j varies linearly in every tetrahedron. Then the space charge density ρ_j at the vertex can be calculated using:

$$\rho_j = \frac{1}{\nu_j} \sum_{i=1}^{N} \omega_i q_i \psi_j(\vec{r}_i), \tag{2.5}$$

where ν_j is the volume of the Voronoi cell at vertex j, and the summation is over all tetrahedra sharing the vertex. An example of the Voronoi cell can be found in [20]. Using this approach, the space charge of a particle is assigned to neighboring vertices according to their relative weights in the linear interpolation functions. Once the potentials at the boundaries and spacecraft components are determined, with the space charge density, Poisson's equation:

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}, \tag{2.6}$$

can be solved using finite elements. Then, the electric field is calculated from the gradient of the potential:

$$E = -\nabla \phi. \tag{2.7}$$

With this expression and the linear (first order) elements used in PTetra, the electric field is constant in a tetrahedron. In general, it is discontinuous at interfaces between adjacent tetrahedra, which would cause inaccuracies when integrating particle trajectories. In order to obtain a continuous electric field across tetrahedra, and more accurate particle trajectories, the electric field $\vec{E}_j$ at each mesh vertex is calculated by averaging the electric fields in the tetrahedra sharing the vertex. The electric field at the particle position in a tetrahedron is then calculated by weighting the electric

fields on the four vertices using the linear interpolation functions:

$$\vec{E}(\vec{r}) = \sum_{j=1}^{4} \psi_j(\vec{r})\vec{E}_j. \tag{2.8}$$

This avoids any discontinuity in the electric field between cells. Once the electric field is determined, the electron and ion positions and velocities are updated using a standard second-order leapfrog algorithm, except when there is a magnetic field. When a magnetic field is present, the electron positions are updated using the more accurate but time-consuming Boris algorithm [24].

PTetra can be used to study various processes related to plasma. For example, the code can simulate the amount of charge collected by each part of a spacecraft. This is particularly interesting since the spacecraft might collect more electrons than ions in space, and be charged negatively with respect to the background plasma. This can cause problems because the performance of some instruments, such as a flowmeter can vary at different floating potentials. PTetra can be used to study such processes. PTetra can also be used to simulate more realistic physical processes than a theoretical model can account for, such as, the accurate dimensions of an instrument, the presence of a ram flow, magnetic fields, photoelectric effects, etc. Instruments such as plasma flowmeters cannot be described using an analytic formula, and inferences made from their measurements must rely on lab calibration or simulation techniques. Such an instrument, however, can be studied straightforwardly using PTetra as described in Chapter 4.

2.2 Machine learning and neural network

Machine learning is the field of study that allows computers to learn from data without being explicitly programmed. There are different types of machine learning techniques: supervised, unsupervised, semi-supervised, and reinforcement learning. One

type of supervised learning is classification, for example, creating a spam filter, where emails are labeled as spam or not spam. Another type of supervised learning, multivariate regression, involves the prediction of a target numerical value. Unlike Newton's first law, there are often many independent variables X corresponding to their dependent variables Y in a given data set. Multivariate regression, which is a method to measure the degree to which more than one independent variable and their corresponding dependent variables are related, can be used to cast the relation between the measured and the inferred parameters in a data set. The feedforward neural network and RBF are special cases of multivariate regression techniques. Contrary to supervised learning, in unsupervised learning, the data are not labeled. Common unsupervised learning techniques include clustering and anomaly detection. In clustering, similar instances are grouped into clusters. Clustering is excellent for customer segmentation, recommender system, search engine, etc. Anomaly detection, on the other hand, learns what "normal" data are, and then tries to detect abnormal instances. For example, outliers located far from other instances in a distribution are likely to be anomalies. Semi-supervised learning has labeled data and unlabeled data. For example, Google Photos might recognize people in a few photos. Once a label is created for a person, that person is labeled in all instances. Reinforcement learning is different from all the above techniques in that it has a rewarding system and learns from mistakes. This is similar to training a mouse to take the right turn to get a piece of cheese, and otherwise receive an electric shock when taking the wrong turn. One example is that the DeepMind's AlphaGo program played the game of Go with itself and trained a model to beat the world champion Ke Jie in May 2017. A good reference for machine learning can be found in [25].

Feedforward neural network

Feedforward neural network is a commonly used supervised learning technique to create multivariate regression models. It contains multiple layers; each consists of

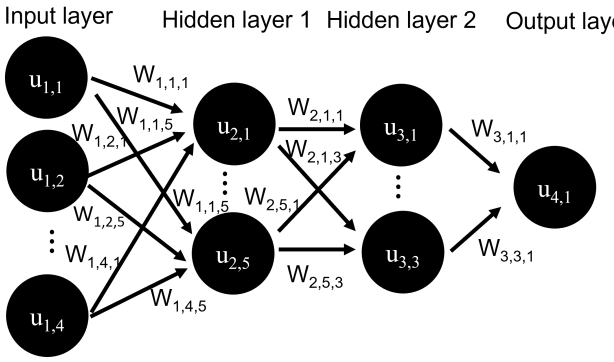

Figure 2.1: Schematic of a feedforward neural network.

various number of nodes. There must be input and output layers, and there can be different numbers of hidden layers between the input and output layers. The input layer contains the independent variable $\bar{X}$ corresponding, for example, to measured probe currents. The output layer contains the numerical value of the inferred value Y, such as the plasma density. Depending on the complexity of the problem, different numbers of hidden layers and nodes can be used in a neural network. The independent variable $\bar{X}$ is a tuple. The dependent variable Y can also be a tuple, but for simplicity, Y is assumed to be a scalar. As shown in Fig. 2.1, each node j in a given layer i in the network is assigned a value $u_{i,j}$, and the node in the next layer $i+1$ are "fed" numerical values from the nodes in the previous layer according to

$$u_{i+1,k} = f\left(\sum_{j=1}^{n_i} w_{i,j,k} u_{i,j} + b_{i,k}\right), \tag{2.9}$$

where $w_{i,j,k}$ are weight factors, $b_{i,j}$ are bias terms, and f is a nonlinear activation function [26]. The network mimics how the human brain operates, thus is called a neural network. The number of hidden layers and the number of nodes in the hidden layers need to be adjusted to fit the specific problem. For example, for larger scattered data with a very complicated structure, more layers and nodes are required. Many activation functions are available to choose from depending on the purpose of the model. For example, a commonly used activation function is the ReLU function defined as $f(x) = \max(0, x)$, more examples of activation function can be found

in [26]. To train a model, a set of data with known independent variable $\bar{X}$, and dependent variable Y need to be provided. Then, the data are fed into a computer program, and the coefficients w and b are initialized with random numbers and then optimized using, for example, the stochastic gradient descent method [25]. A trained model contains optimized coefficients w and b. For a given independent variable $\bar{X}$ in the input layer, the result from the output layer needs to be close to the known dependent variable Y. In order to assess the performance of a model, a cost function (also called loss function) needs to be defined.

Cost function

A cost function has the following properties: i) it is non negative, ii) it vanishes if model inferences agree exactly with a given data set, and iii) it increases as inferences deviate from actual data. For example, the root mean square error is defined as:

$$RMS = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} (Y_{mod_i} - Y_{data_i})^2}, \tag{2.10}$$

where Y_{data} and Y_{mod} represent respectively the known and inferred plasma parameters, and N_{data} is the total number of data points. Thus, the goal of training a model is to minimize the cost function while varying the various coefficients. The choice of the cost function depends on the parameter to be inferred. Relative errors are better for parameters which never vanish, and vary over several orders of magnitudes, as for example, the density. Absolute errors, on the other hand, are better for parameters which do not vary as much, or which may be close to zero, as for example, the satellite floating potential. The choice of the cost function also depends on the nature of the parameter. For example, the mean squared error is a good choice when the noise in the training data has a normal, or Gaussian distribution, while the maximum absolute error provides the most conservative error estimates on inferences when the distribution of measurement errors is unknown. Feedforward neural network models in this thesis are created using Tensorflow from Google [27]. Applications of neural

networks can be found in Chapter 4 and 5.

2.3 Radial Basis Function

The RBF models in this thesis are created using the CERBF program [1]. The basis of RBF regression is straightforward, the dependent variable Y is given by:

$$Y = \sum_{i=1}^{N} a_i G\left(\|\bar{X} - \bar{X}_i\|\right), \tag{2.11}$$

where a_i are interpolation coefficients, G is the interpolating function, $\bar{X}$ are the independent variables, and $\bar{X}_i$ are the centers. The Euclidean distance between n-tuples $\bar{P}$ and $\bar{Q}$ is defined:

$$\|\bar{P} - \bar{Q}\| = \left(\sum_{i=1}^{n} (p_i - q_i)^2\right)^{1/2}. \tag{2.12}$$

where p_i and q_i are respectively the i-th components of n-tuples $\bar{P}$ and $\bar{Q}$. The G function has radial symmetry and only depends on the Euclidean distance; whence the name "radial" in "Radial Basis Function". Illustrations of the RBF regression approach are given in Fig. 3.1 and 3.2, this section will focus on other aspects of RBF. There are four important factors to consider when training an RBF model, they are the cost function, the number of centers, the G function, and the distribution of centers. The cost function has been discussed in the previous section.

Number of centers

In RBF, the number of centers determines the complexity of a model. Too complex a model will result in overfitting, as illustrated in Fig 2.2. The full data set to construct a model is split into two disjoint sets, the training set, and validation set. Once the model is trained, it is applied to the validation set to calculate the validation error without further optimization. There are certain limitations on the refinement of a model on a training set, such that further improvement of model inference skill in

[1] The CERBF program can be found at: https://codeocean.com/capsule/3524673/tree

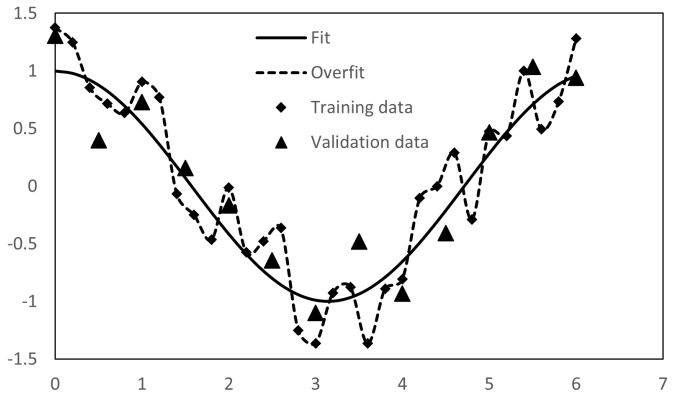

Figure 2.2: Illustration of overfitting.

the training set should result in reduced model inference skill in the validation set. A good model is one with the right level of complexity, which fits the training data and validation data roughly equally well. For RBF, the complexity of the model increases with the number of centers, and so does the computational cost. Thus, the right number of centers needs to be determined carefully. In practice, the optimal number of centers can be determined by gradually increasing the number until the validation cost function reaches a minimum. A special case used to create models for large scattered data is the Radial Basis Function partition of unity (RBF-PU) method [28]. In this approach, essentially all entries in the training set are used as centers, and the RBF interpolants are constructed in local sub-domains, which are then combined to form the global interpolant. In this case, the training error is zero, thus the validation error from the validation set is especially important as it gives a measure of the inference skill of the model.

RBF G function

A G function must be selected a-priori to train an RBF model. For a specific problem, many G functions are tested, and the one selected is the one which produces the lowest cost function. The G function can be global as $G(r) = r$ or local as $G(r) = e^{-r^2}$. A global G function generally vanishes at the centers where they are defined and it is

significantly different from zero away from the center. In contrast, a local G function is significantly different from zero only near the center. Different G functions have different shapes in the parameter space and are used for different purposes. For example, only local G functions are used in the RBF-PU method; otherwise, the matrix involved is not sparse and is hard to solve. A sparse matrix means that most of the matrix elements are zeros. In general, only a fraction of the nodes in the training data set are used as centers, and there will be a training error and a validation error. In this case, due to the smaller dimension of the matrix, it is more practical to use global G functions. The optimal G function depends on the specific problem to solve. Essentially, given a G function, training an RBF model is to find the right centers and the interpolation coefficients that lead to the lowest cost function for the validation set.

Distribution of centers

Center positions are generally selected initially from the training data set. If the training data contains $\mathcal{N}$ nodes, and N centers are used to approximate the RBF function, the most straightforward exhaustive approach to select N centers among $\mathcal{N}$ nodes is to consider all combinations of N centers among the $\mathcal{N}$ possible nodes. The total possible number of choices is given by:

$$\binom{\mathcal{N}}{N} = \frac{\mathcal{N}!}{N!(\mathcal{N}-N)!}. \tag{2.13}$$

For every combination of centers, the interpolation coefficients a_i must be solved using:

$$\sum_{i=1}^{N} a_i G(\|\bar{X}_k - \bar{X}_i\|) = Y_k, \tag{2.14}$$

for $k = 1,\ ...,\ N$. This can be written in matrix form as:

$$\begin{pmatrix} G(\|\bar{X}_1 - \bar{X}_1\|) & \cdots & G(\|\bar{X}_1 - \bar{X}_N\|) \\ G(\|\bar{X}_2 - \bar{X}_1\|) & \cdots & G(\|\bar{X}_2 - \bar{X}_N\|) \\ \vdots & \ddots & \vdots \\ G(\|\bar{X}_N - \bar{X}_1\|) & \cdots & G(\|\bar{X}_N - \bar{X}_N\|) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}. \tag{2.15}$$

The independent variables $\bar{X}$ and the dependent variables Y are known from the training data set, thus the interpolation coefficients a_i can be determined using methods such as Gaussian elimination. With the coefficients and Eq. 2.11, we are ready to solve for the dependent variable Y_{mod} for each of the independent variable $\bar{X}$ in the training set. Then with the known values of dependent variables Y_{data}, the cost function can be determined using, for example, Eq. 2.10. The combination of centers corresponding to the lowest cost function and their interpolation coefficients will be the outcome of the training. On the other hand, the optimal position of the centers may be close to, but not necessarily exactly at the nodes in the set itself. The model can be further optimized by considering the "nugget effect" [29] as described in Chapter 3.

CERBF approach

The problem with large training data set is that the number of possible combinations can be prohibitively large. The CERBF program uses a new original algorithm that is particularly efficient in determining centers. The strategy consists of successively carrying out straightforward exhaustive searches on randomly selected small subsets from the training set while calculating the cost function over the full training set. In each step, the optimal centers found in the previous iteration are carried forward to the next iteration, and the model is optimized by accounting for the "nugget effect". With a few hundreds of fast iterations, CERBF can often find a model with a lower cost function than with the much more time-consuming exhaustive search. The algorithms, applications, and user's guide are described in detail in Chapter 3.

The program and the implementation of the Message Passing Interface (MPI) in the program are briefly discussed in Appendix A.

Chapter 3

Inference of m-NLP data using Radial Basis Function regression with center-evolving algorithm

This chapter is based on a paper published in Computer Physics Communication titled "Inference of m-NLP data using Radial Basis Function regression with center-evolving algorithm" by Guangdong Liu, and Richard Marchand. The authors thank A.Olowookere for testing the algorithm. This work was financially supported by the China Scholarship Council (CSC), the Natural Sciences and Engineering Research Council of Canada, and Compute Canada.

3.1 Introduction

Langmuir probes are useful instruments on spacecraft or in lab plasma experiments to measure plasma parameters such as density, temperature, and spacecraft potentials. A Langmuir probe consists of an electrically biased conductor immersed in a plasma from which current is collected and measured. Probe characteristics; that is, collected currents as a function of probe bias voltages, can then be used to infer plasma and satellite parameters. The inference of these parameters is usually based on analytic theories, which leads to analytic expressions for collected currents as a function of probe voltages, and which can be implemented with modest computational resources. A very common model used in this context is the Orbital Motion Limited (OML)

theory developed nearly a century ago by Langmuir and co-workers in the 1920s [8]. However, a difficulty with analytic expressions derived from theory, is that they are severely limited, owing to the many assumptions made, for these theories to lead to analytically tractable solutions. For example, with cylindrical probes, the OML theory assumes Debye lengths much larger than the probe radius, currents collected per unit length identical to those collected by an infinitely long probe, and the theory neglects the presence of any other nearby object. The use of such theories under conditions in which all underlying assumptions are not well satisfied can lead to uncontrolled uncertainties in inferred physical parameters. This predicament has motivated several recent studies aimed at finding inference techniques to account for conditions and physical processes unaccounted for in theoretical models. For example, Chalaturnyk and Marchand [30] showed that RBF-trained models could accurately reproduce plasma densities and temperatures from spherical probe characteristics calculated in the OML approximation. Similarly, Guthrie, et al. [31] constructed, and assessed the skill of RBF regression models from synthetic data, applied to fixed bias multiple needle Langmuir probes (m-NLP). Their approach was then applied to in situ data from the Visions-2 mission, with an excellent qualitative and quantitative agreement between their model inferences, and reported values. Models have also been developed and assessed, using synthetic data for fix bias spherical probes, to infer a satellite potential [32], plasma density, and the ratio between the plasma density and the square root of the temperature [32]. These previous studies made use of the RBF regression technique with relatively small training data sets in which a straightforward search of the optimal set of inference centers was practical. With larger data sets, however, the time required to do a straightforward exhaustive search increases very rapidly with the size of the training set, and the number of centers considered. In this work, three-dimensional particle in cell simulations are used to compute the currents collected by the m-NLP in order to build a solution library which is then used to create multivariate regression models to infer plasma density. The size of the resulting

solution library in this, and many similar problems, is such that an exhaustive search would be too time consuming to be applicable in practice. The evolutive search strategy described in this paper provides an efficient solution to this problem, by significantly reducing the time needed to select near-optimal regression centers, while allowing for the construction of models with inference accuracy comparable to those obtained with straightforward exhaustive searches.

In complex systems where the relation between multiple independent variables and dependent variables cannot be cast analytically with theory, a useful alternative is to construct approximate predictive or inference models based on multivariate regression. Several approaches have been proposed to achieve this objective, including collocating analytic fits on selected data points or centers, kriging, radial basis functions, or neural networks. The radial basis function considered here is particularly convenient because of its relative simplicity and good accuracy. Kriging developed by D. Krig in the 1950s was the first regression approach used to interpolate a dependent variable (mineral contents) from scattered samples [33]. Kriging and RBF are nearly the same; the main difference being that kriging requires the solution of a system of linear equations of size $N + d$ by $N + d$, where N is the number of centers, and d is the dimension, or more precisely, the number of independent variables in each entry $\bar{X}$ in the data set. Conversely, with RBF, the linear matrix equation which needs to be solved is only of size N by N. Thus, other than the additional d rows in kriging needed to remove the "bias" in the inference model, the two methods are the same. The radial basis function method as it is known today was developed in 1971 by Hardy [34], who introduced the multiquadric method for topological applications. Since then, much work has been reported to extend and optimize RBF in speed and accuracy [28, 35–43]. The major difficulty, when applying multivariate RBF to data in which there is no analytic relation between multi-dimensional independent variables $\bar{X}$ and dependent variables Y, is that N centers have to be selected at, or close to nodes $\{(\bar{X}_i, Y_i)\}$ in a training set of size $\mathcal{N}$. The straightforward, or exhaustive strat-

egy, consisting of constructing models for all possible combinations of $\mathcal{N}$-choose-N centers in a training set, can be prohibitively time consuming when $\mathcal{N}$ is large. Other strategies have been developed, such as k-clustering [38, 42], and Gaussian clustering [40, 41]. While these center distribution strategies are significantly faster than the straightforward exhaustive search, they have disadvantages of i) requiring many more centers, and ii) not accounting for the "topography" of the data set in finding the best distribution of a given number of centers. Excellent reviews of RBF have been given in the literature, as in [44–47] where the reader can find more detailed descriptions. In this paper, we introduce another algorithm, the Center-Evolving (CE) algorithm, and identify its advantages with examples.

The basic concepts of RBF are reviewed in Sec. 3.2, where our new center evolving approach is described. Example results are presented in Sec. 3.3, with quantitative assessments of model accuracies. A guide for using the code is presented in Sec. 3.4, and a summary of our findidngs and some concluding remarks are in Sec. 5.5.

3.2 Methodology

In this section, we describe the general RBF approach, in which multivariate regression is obtained from a linear superposition of functions of the L^2 norm (Euclidean distance) between an arbitrary point and reference points, or centers, in a multivariate space of independent variables. This is followed by a description of a straightforward exhaustive search for the best centers; that is, the set of N centers which minimizes a selected cost function on a given training set. We then present a modified multi-step search strategy based on an adaptation of the exhaustive search, using subsets of the full training set.

3.2.1 The RBF regression approach

Given a data set consisting of independent input n-tuples $\bar{X} = (x_1, ..., x_n)$ with corresponding dependent output Y, RBF regression of Y on $\bar{X}$ is written as

$$Y = \sum_{i=1}^{N} a_i G\left(\|\bar{X} - \bar{X}_i\|\right). \tag{3.1}$$

In general, Y could also be a tuple, but for simplicity, and without loss of generality, we limit our attention to scalar dependent variables. In Eq. 3.1, the $\bar{X}_i$ represent the N centers, G is the interpolating function, and the a_i are interpolation coefficients which can be determined by requiring collocation at the centers, by solving the system of linear equations

$$\sum_{i=1}^{N} a_i G(\|\bar{X}_k - \bar{X}_i\|) = Y_k \tag{3.2}$$

for $k = 1, ..., N$. Improvements to this prescription will be discussed later, but for now, we assume simple collocation as a first approximation. With this expression, RBF regression is seen to involve a linear superposition of a function G of distances between centers $\bar{X}_i$, and $\bar{X}$, as illustrated schematically in Fig. 3.1 for a two-dimensional space of independent variables.

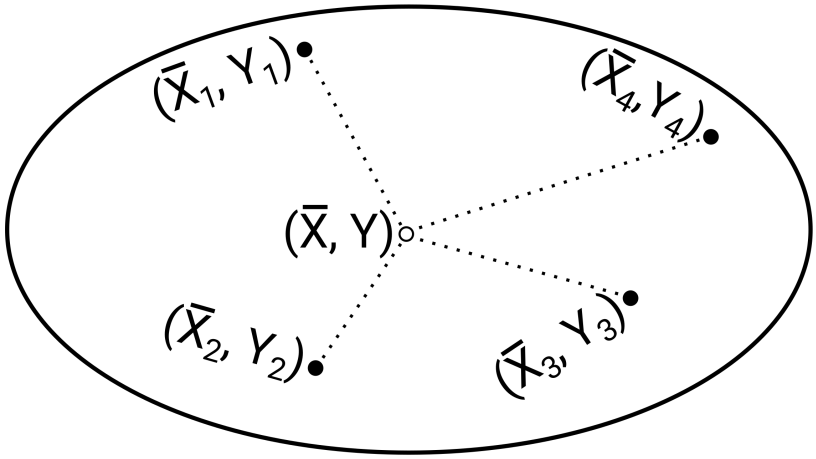

Figure 3.1: Illustration of centers in a two-dimensional space of independent variables, from which a dependent variable Y at a point $\bar{X}$ is inferred.

To be specific, it is customary to define the "distance" between two n-tuples as

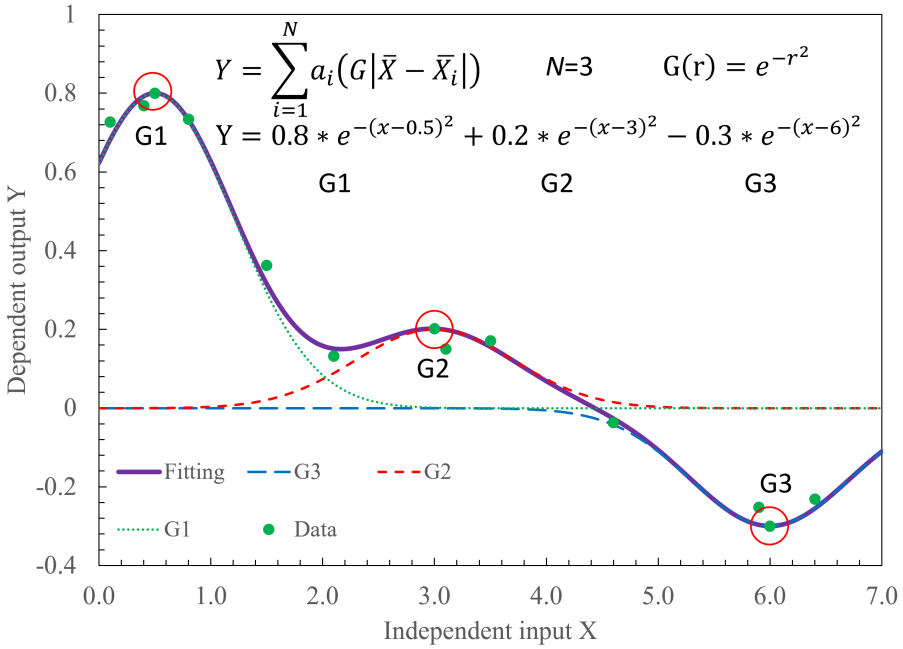

Figure 3.2: Illustration of localized G functions in one dimension. Here the coefficients a_i are determined by requiring collocation at the centers (red circles).

their Euclidean distance, or the L^2 norm of their difference; that is

$$\|\bar{X} - \bar{X}_i\| = \left(\sum_{j=1}^{n}(x_j - x_{i,j})^2\right)^{1/2}, \tag{3.3}$$

where x_j and $x_{i,j}$ are respectively the j-th components of n-tuples $\bar{X}$ and $\bar{X}_j$. The G function has radial symmetry and only depends on the Euclidean distance; whence the name "radial" in "Radial Basis Function". The interpolating function G appearing in Eq. 3.2 can be chosen arbitrarily provided that the system of equations be non-singular. Two types of interpolations can be used, consisting of "localized" functions, which are significantly different from zero, only close to the centers, and "global" functions, which generally vanish at the nodes where they are defined and are non zero over the remaining range in $\bar{X}$ parameter space. With localized G functions, the calculation of Y at a given n-tuple $\bar{X}$ is determined mostly from nearest neighbors in the set of $\bar{X}_i$ centers. This is illustrated schematically in Fig. 3.2 for a one-dimensional space of independent variables. With global functions, on the other hand, Y is determined from a combination of all centers. With local G functions, fast

Table 3.1: Example local and global G functions used in RBF regression. In each case, c is an adjustable parameter which can be set to optimize accuracy.

G function	Equation ($r = \|\bar{X} - \bar{X}_i\|$)	Type
Gaussian	$G(r) = e^{-cr^2}$	local
Inverse quadratic	$G(r) = \frac{1}{1+(cr)^2}$	local
Multiquadric	$G(r) = \sqrt{1+(cr)^2}$	global
Polyharmonic spline	$G(r) = r^k$, k=1,3,5,...	global
Polyharmonic spline	$G(r) = r^k \ln(r)$, k=2,4,6,...	global

efficient algorithms exist to distribute large numbers of centers in which case, regressions are effectively made from nearest neighboring centers. This in turn requires a sufficiently large number of centers, and dense distribution of these centers, such that there is overlap in the G functions between near neighbors. In this case, the exact number and distribution of centers are less critical to the accuracy of the model, as long as every region of interest in parameter space is "within range" of a few centers. If, on the other hand, regression is made in a region far from any nearby center; that is, where $G(|\bar{X} - \bar{X}_{nn}|)$ is zero or very small, where $\bar{X}_{nn}$ is a representative nearest neighbor, then the accuracy of the interpolation would likely be compromised. For models constructed with few centers, however, global interpolation functions are preferred, as every point in parameter space would be in the range of every center. In this case, the choice of G and the exact distribution of the few centers are critical in order to obtain good accuracy. This is considered in what follows, but for reference, examples of G functions of both types are given in Table 3.1.

3.2.2 The cost function

The assessment of a model predictive or inference skill relies on a cost function (also called loss function) with the following properties: i) it is non negative, ii) it vanishes if model inferences agree exactly with given data in a given set, and iii) it increases

as inferences deviate from actual data. Several possibilities exist for such functions, including the root mean square error,

$$RMS = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} (Y_{mod} - Y_{data})^2}, \tag{3.4}$$

the root mean square relative error

$$RMSr = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} \frac{(Y_{mod} - Y_{data})^2}{Y_{mod}^2}}, \tag{3.5}$$

the maximum absolute error

$$MAE = \max \{|Y_{mod} - Y_{data}|\} \tag{3.6}$$

and the maximum absolute relative error

$$MRE = \max \left\{ \left| \frac{Y_{mod} - Y_{data}}{Y_{mod}} \right| \right\}, \tag{3.7}$$

where Y_{data} and Y_{mod} represent respectively known and inferred dependent variables, and N_{data} is the total number of data points. Many more cost functions can of course be constructed, for different types of applications, provided that they satisfy the three criteria stated above. The four cost functions defined in 3.4 to 3.7 are implemented in `CERBF`, and can be selected conveniently from the code input file.

3.2.3 Exhaustive search RBF

In order to perform an exhaustive search on a training data set with $\mathcal{N}$ nodes, models must be constructed for every possible combination of centers. In each case, the system of Eq. 3.2 has to be solved for the interpolation coefficients a_i, and the cost function must be evaluated over the full training set. The optimal model is then selected as the one with the best inference skill; that is, the one that yields the smallest value of the cost function. For large training sets, the total number of combinations

$$\binom{\mathcal{N}}{N} = \frac{\mathcal{N}!}{N!(\mathcal{N} - N)!}, \tag{3.8}$$

can be prohibitively large, which motivated the development of the more efficient technique presented below.

3.2.4 Center evolving algorithm

The Center Evolving (CE) algorithm presented here provides an efficient and accurate alternative to the exhaustive search described in 3.2.3. The strategy consists of successively carrying out exhaustive searches on randomly selected small subsets of size $\tilde{\mathcal{N}}$, from the full $\mathcal{N}$ element training set, while calculating the cost function over the entire training set. Finding an optimal set of centers on any given such subset will take much less time than performing an exhaustive search on the full training set. A model constructed from centers found from such a single subset would generally have low inference skill. The reason is that a single random sample is unlikely to contain the optimal set of centers, which would be found in an exhaustive search. However, convergence to an optimal distribution of centers is dramatically improved by adopting an evolutive strategy whereby, optimal centers found in one subset, are kept in the following one. To be specific, assuming that the optimal centers found with the first randomly selected subset are $\{X_1, ..., X_N\}$, then these centers are included in the next subset, or next iteration, in which only $\tilde{\mathcal{N}} - N$ entries are actually selected randomly from the full training set, excluding of course the optimal centers found thus far. In this process, it is important to select additional nodes while excluding the centers retained from the previous iteration, because if some of these centers were selected again in an unconstrained random selection, the occurrence of two or more identical nodes in the training subset would lead to a singular matrix in Eq. 3.2. An exhaustive search is then carried out on the resulting subset involving all combinations of nodes, including the optimal ones from the previous subset. This typically results in a new set of optimal centers; with some previous centers remaining as optimal, and others being replaced by new ones. With this strategy, the cost function calculated from one subset to the next never increases, convergence is relatively fast, and the resulting

model skills are found to be very close to those obtained with an exhaustive search on the full training set.

The CE approach has been found to be both fast and to produce models with good inference skills with several large training sets. For example, a CE search with a batch size of 20 using 5 centers in a 200 entry training set typically requires 600 times less computer time than an exhaustive search on the full training set, with a modest (few $) gain of accuracy. A significant advantage of the CE algorithm in practice, is that the computational cost can be controlled by selecting the size of the subset $\tilde{\mathcal{N}}$ and the number of iterations involved. With this gain in efficiency, it is also possible to construct models with more centers, and thus more inference skills.

3.2.5 Parallel computing

An obvious advantage of using parallel processing is an increase in speed. The CERBF is written so as to run on multiple processors in parallel using the Message Passing Interface (MPI) [48]. In an exhaustive search, most of the computation time is spent on going through all the possible combinations. The use of parallel processing is particularly efficient due to the fact that each processor considers a distinct subset of the possible combinations so that this work is split between multiple parallel processors. Communications with the main process only occur at the start and end of each CE iteration, and it involves small data sets, consisting of the selected training subset $\tilde{\mathcal{N}}$, optimal centers, and cost functions.

3.2.6 The nugget effect

A further improvement in training accuracy is obtained by accounting for the "nugget effect", as it is referred to in geostatistics [29]. This is motivated by the fact that the optimal position of the centers for a given training set may be close to, but not necessarily exactly at nodes in the set itself. Thus at the end of each iteration in

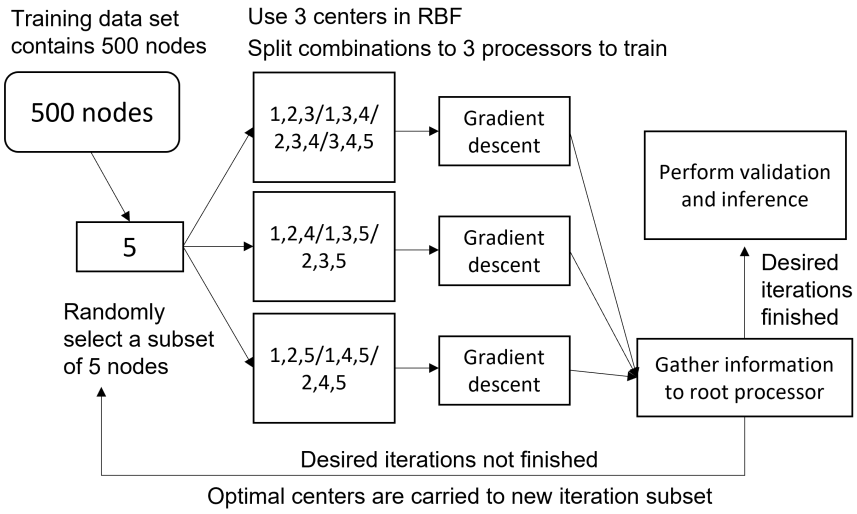

Figure 3.3: Schematic illustration of CERBF. In this particular example, the training data set has 500 nodes, and the CE iteration batch size $\tilde{\mathcal{N}}$ is 5. The RBF model uses 3 centers, and the workloads of 5 choose 3 are distributed between three processors, where Eq. 3.2 are solved, and the cost functions are calculated over the 500 nodes. Variations in the center points, when applying nugget corrections, are made with gradient descent, so as to minimize the cost function.

the CE strategy, given a set of optimal centers found in a given processor, deviations are considered in the vicinity of these centers and in the values of the corresponding dependent variables

$$\left\{\bar{X}_i, Y_i\right\} \rightarrow \left\{\bar{X}_i + \delta\bar{X}_i, Y_i + \delta Y_i\right\}, \tag{3.9}$$

and the cost function is further minimized relative to these deviations. In the code, this optimization is made with straightforward gradient descent, which can lead to an appreciable (several %) increase in inference accuracy.

3.2.7 CERBF schematic

In summary, the CE algorithm consists of the following steps:

1. Select a cost function and an interpolating function G.

2. Specify the number of centers N, the size of the subset $\tilde{\mathcal{N}}$ in each CE training iteration.

3. Select $\tilde{\mathcal{N}}$ elements randomly from the full training set.

4. Determine optimal centers among subsets of the $\tilde{\mathcal{N}}$ randomly selected elements. This is done while keeping optimal centers from one iteration to the next, as described in 3.2.4 to 3.2.6.

5. Repeat step 3 and step 4 for a specified number of iterations.

A schematic illustration of CERBF can be found in Fig. 3.3. For RBF, the model complexity is proportional to the number of RBF centers. Increasing the number of centers will result in a better inference skill in the training data set, but beyond a certain number, overfitting generally occurs. Thus the full data set is split into two distinct sets: the training set and the validation set. Model construction is done only on the training set by minimizing the cost function on that full set. The validation error is calculated by applying the model to the validation set without further optimization. The data will be overfitted if a further increase of RBF centers increases the inference skill in the training set, while leading to an increase in the cost function evaluated on the validation set.

3.3 Example results

The CERBF program is written in Fortran 90 with the Message Passing Interface (MPI) [48]. The m-NLP synthetic data set is created using the particle in cell code PTetra [19, 20]. The independent variables $\bar{X}$ consist of currents collected by four needle probes, each one being biased to a different voltage V_b relative to a satellite, at various satellite potentials V_f, and for electron densities and temperatures in a range relevant to ionospheric plasma. The probe bias voltages and the range of satellite potentials are such that, probe potentials relative to the background plasma, $V = V_f + V_b$ are in the electron saturation regime. These parameters are summarized in Table 3.2. The dependent variable Y for which an inference model is constructed is the electron density. Training is done with the known densities assumed in the simulations, and

inference model skills are assessed by comparing model inferences with known densities in our synthetic data set. Doing an exhaustive search as described in Sec. 3.2.3 on a training data size of 200 nodes with 5 centers would require 2.5×10^9 combinations. Using $G(r) = r^3$, gfortran, and the Open MPI implementation of the message passing interface, it takes approximately 9.5 hours to complete the calculation on six hyperthreads on an AMD Ryzen 5 2400G processor. The resulting maximum relative error of the model, when applied to the full training set is 15.7%. The correlation plot in Fig. 3.4 shows the inferred densities as a function of actual densities used in the simulations.

The same problem was solved, under the same conditions, but using the CE technique with subsets consisting of $\tilde{\mathcal{N}} = 20$ nodes, and 200 CE iterations. The calculation completed in approximately 55 seconds; resulting in a maximum relative error of 13.1 %. With a training set of 200 nodes, the time required to train a model with CE iterations is smaller than that required with an exhaustive search, by more than two orders of magnitude. Interestingly, this gain in efficiency is also accompanied by a slight gain in inference accuracy. The increase in accuracy may be surprising, owing to the fact that not all possible combinations of centers are considered. The explanation is that many more optimizations are made and compared to one another, in which nugget corrections are applied, and we find empirically that the best models found after applying this correction, are not necessarily the ones which would produce the lowest cost function without the nugget correction.

With larger training sets, the gain in computing speed would be even more significant. More tests were made using a synthetic training data set consisting of 3000 nodes which would require 2.0×10^{15} combinations if an extensive search were performed. Based on the calculation time of the extensive search on 200 nodes on six hyperthreads, the estimated time to complete this calculation would be about nine hundred years. Under the same conditions, with a batchsize of 20 using 5 centers,

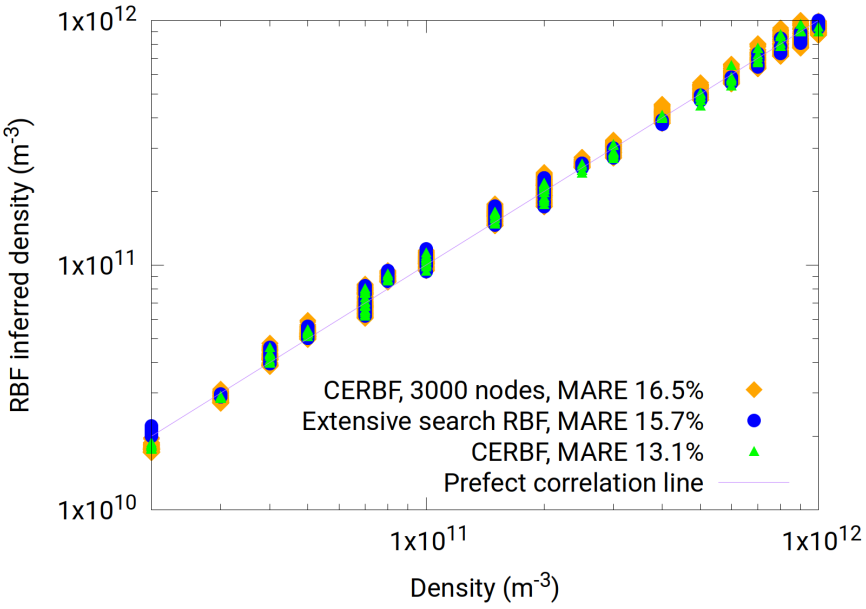

Figure 3.4: Correlation plot comparing CERBF training on 3000 nodes (16.5%), Extensive search RBF (15.7%) and CERBF search (13.1%), both on 200 nodes. RBF regression is used to infer plasma densities from collected currents in a synthetic data set.

Table 3.2: Plasma conditions considered when constructing synthetic data.

Name	Unit	Range
Electron density ne	m^{-3}	$2 \times 10^{10} - 1 \times 10^{12}$
Floating potential V_f	V	$-1 - -6$
Bias potential V_b	V	6 8 9 10
Electron temperature T_e	eV	$0.07 - 0.3$

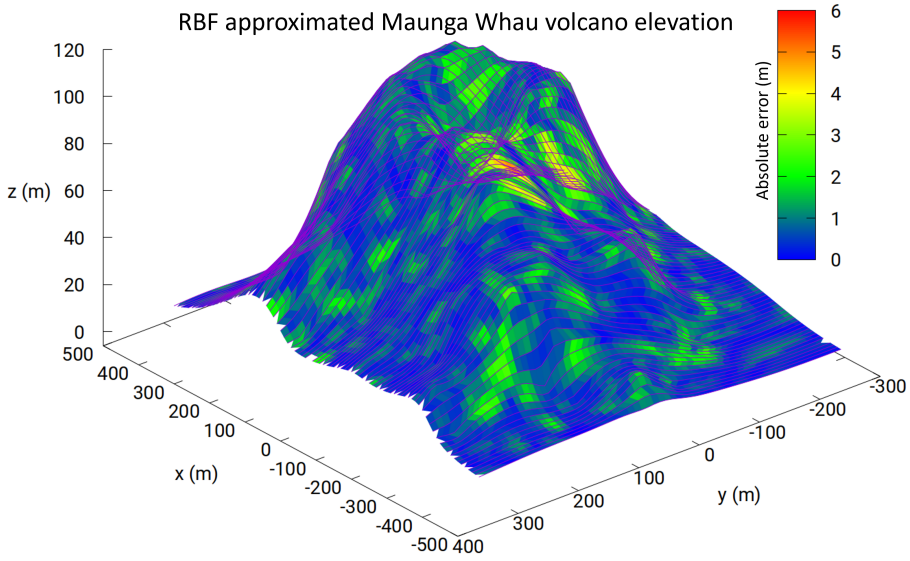

Figure 3.5: Illustration of CERBF approximated Maunga Whau volcano elevation. The RBF approximation uses 60 centers with $G(r) = r$. The color bar in the plot refers to the absolute error between the RBF approximation and the measured data. The training and validation RMS errors are both 1.2 meters.

200 CE iterations require 3.1×10^6 combinations which finishes in 8 minutes, with a maximum relative error of 16.5 % as shown in Fig. 3.4. This demonstrates that the CERBF is a highly efficient tool to create accurate RBF models for large-size data.

Another application related to Earth's topography is also considered. This example has been studied by Cavoretto and co-workers as an interpolation problem using the Radial basis function partition of unity (RBF-PU) method [28]. In this approach, all data in the training set are used as centers, and the RBF interpolants are constructed in local sub-domains, which are then combined to form the global interpolant. The data set contains 5307 volcano elevation measurements from Maunga Whau (Mt. Eden) in Auckland, New Zealand [49]. Cavoretto uses 5200 entries as the training data set, and the rest of the 107 entries as the validation data set. The same data are approximated using the CERBF approach with 5200 randomly selected entries in the training set and the other 107 entries as the validation set. In this study, we first normalized the elevation data by subtracting all elevations by the lowest elevation,

thus the elevation will start from zero meter. Using $G(r) = r$ with 30, 40, 50, and 60 centers, the training RMS errors are 1.8, 1.5, 1.2, and 1.2 meters respectively. The RBF approximated volcano Maunga Whau elevation using 60 centers is shown in Fig. 3.5, the model is trained in three hours using 12 cores on the Compute Canada Cedar cluster with Intel E5-2650 V4 Broadwell 2.2 GHz CPU. In each case, validation is done by applying the trained models to the validation set, which results in RMS errors of 1.9, 1.6, 1.4 and 1.2 meters, respectively. The validation error reported by Cavoretto is in the range of 0.7 to 1.1 [28], which is close to the validation error from our method using 60 centers. Interestingly, although interpolation problems can be particularly computational costly using global RBF interpolation functions for large scattered data, our approach works relatively well with a global interpolation function, as illustrated in this example. The advantage of using CERBF is, of course, much fewer centers are required to interpret the data, e.g. 60 vs 5200 centers in this example. With fewer centers, the computational cost of inferences will be lower, and less storage will be required.

3.4 User's guide for program CERBF

The `CERBF` software package consists of a Fortran program, a makefile, a readme file, input and output files [1]. In this section, we briefly describe these components and provide user's guide to program CERBF.

3.4.1 Input files

Two input files are required for training. The file '`rbf.dat`' contains all the parameters needed to construct and apply a model, including the names of all other files involved. This is the only file with a fixed name. All other files can be referred to by name, with absolute or relative paths, in '`rbf.dat`'. Input variables used in the code are all part of namelists, which allows the user to specify default values in the

[1] The CERBF program can be found at: https://codeocean.com/capsule/3524673/tree

Table 3.3: Variables appearing in `rbf.dat`, used to control the execution of CERBF.

`ncenter`	number of centers in training
`batchsize`	size of subset $\tilde{\mathcal{N}}$ in CERBF
`nbatch`	maximum number of iterations in CERBF
`xdim`	dimension of independent variable X
`ydim`	dimension of dependent variable Y
`costfcntype`	type of cost function
`trainingset`	name of the training data file
`validationset`	name of the validation data file
`inferenceset`	name of the inference data file
`ndescent`	maximum number of gradient descent iteration
`lr`	learning rate in gradient descent
`shownug`	used to show gradient descent diagnostics
`depind`	specifies columns of Y to use in training

program, or to specify different possible values for a given variable, using multiple lines; the last occurrence of a given variable being the one used in the program. Each variable is briefly described in the readme file, but for reference, Table 3.3 gives a summary of the variables, with their meaning.

In the training data file, the actual training data begins after '`$Begin data`', followed by '`ndata=`' to specify the number of data points to be used. From left to right, the data file should contain first the $\bar{X}$ columns, then the Y columns. The choice of the G function needs to be specified in '`CERBF.f90`' under function '`g`', and of course, the program has to be compiled each time the source code is modified. The types of cost functions can be specified in the '`rbf.dat`' file. The learning rate '`lr`' which is related to the step size in gradient descent needs to be adjusted to fit the specific problem. A small learning rate will slow convergence, too large a learning rate will lead to failure in the minimization of the cost function.

The program can also do validation and inference, by setting the validation and inference file names following '`validationset=`' and '`inferenceset=`' respectively. If these variables are set to '`null`', then the program will not do validation or inference. It is important to use the same G function, and number of centers when validating and inferring, as used when training a model. The validation and inference input files should have the same format as the training input file, except for the Y columns, which are not required for the inference file. Training, validation, and inference can be done in a single execution. The program can also train and validate, train and infer, only validate, or only infer in a single execution, as specified in the input file '`rbf.dat`'. The dimensions of $\bar{X}$ and Y need to be specified in '`rbf.dat`' with '`xdim=`' and '`ydim=`'. The dependent variable Y can be multi-dimensional in the training data, and users can choose to train on all the columns of Y or select one or a few columns to train. For example, if the user wants to train on the second and third components of Y and ignore the first column of Y, then the user can set '`ydim=2`', and '`depind=2 3`'. Several other parameters appear in the '`rbf.dat`' file where they are briefly described.

3.4.2 Output

The output of the program depends on the options selected by the user. When training, the program will produce a '`centers.out`' file containing the best model found so far in the run. This file is required for validation or inference. The program can be stopped gracefully by creating a file named '`.quit`' in the folder where it is running. Training can later be restarted from the '`centers.out`' file if this file exists in the current directory. The program produces different '`training.out`' and '`validation.out`' files depending on the dimension of Y. If Y has dimension d, the first d columns in these files, contain the known values of Y from the training or validation file respectively, followed by d columns of the inferred Y. The last d columns contain the differences between the inferred and known values of Y. When

the code is used in inference mode, the first '`xdim`' columns in the output file contain the data $\bar{X}$ from which inferences are to be made, followed by d-tuples of the inferences Y.

3.5 Summary and conclusion

A computer code CERBF is described which uses a new algorithm, the Center Evolving (CE) algorithm, to select near-optimal centers in RBF regression to make predictive models to infer plasma conditions from m-NLP data. The CERBF is designed to be used on large training data sets. It is much faster than a straightforward exhaustive search RBF algorithm by sequentially training models on small subsets of the training data, while carrying the best set of centers found in one subset, into the next, so as to progressively improve inference skills from one iteration to the next. The main advantage of CERBF compared to the straightforward exhaustive approach is that, for practically the same accuracy, it converges significantly faster on large training data sets. Also, the required computational resources scale more favorably with the number of centers used in a model. In tests made with a training set consisting of 200 nodes, models with comparable inference skills can be constructed faster by more than two orders of magnitude, using CERBF, compared to the exhaustive approach. This shows that the proposed CERBF approach is a promising technique to efficiently construct regression inference models for large data sets.

Chapter 4

Kinetic simulation of segmented plasma flow meter response in the ionospheric plasma

This chapter is based on a paper published in the Journal of geophysical research: space physics, entitled "Kinetic simulation of segmented plasma flow meter response in the ionospheric plasma", by Guangdong Liu, and Richard Marchand. This work was supported by the China Scholarship Council (CSC) and the Natural Sciences and Engineering Research Council of Canada. The kinetic simulations used in this study were made on the Compute Canada computing infrastructure. Simulation data can be accessed through https://zenodo.org/record/4434879#.X_4BAdhKhPZ [50].

4.1 Introduction

Plasma winds are a key manifestation of the dynamical processes at play in the ionosphere, including ionospheric coupling with the magnetosphere and with solid Earth. This has motivated the use of various instruments mounted on satellites to measure plasma flow velocities under different space plasma environments. In addition to affecting ground infrastructures [51], events such as magnetic storms or substorms can be responsible for satellite malfunction and, in extreme cases, total loss [52]. These storms cause turbulence in the magnetosphere, which often result in strong currents and winds. Thus monitoring ionospheric winds provides key information for a better

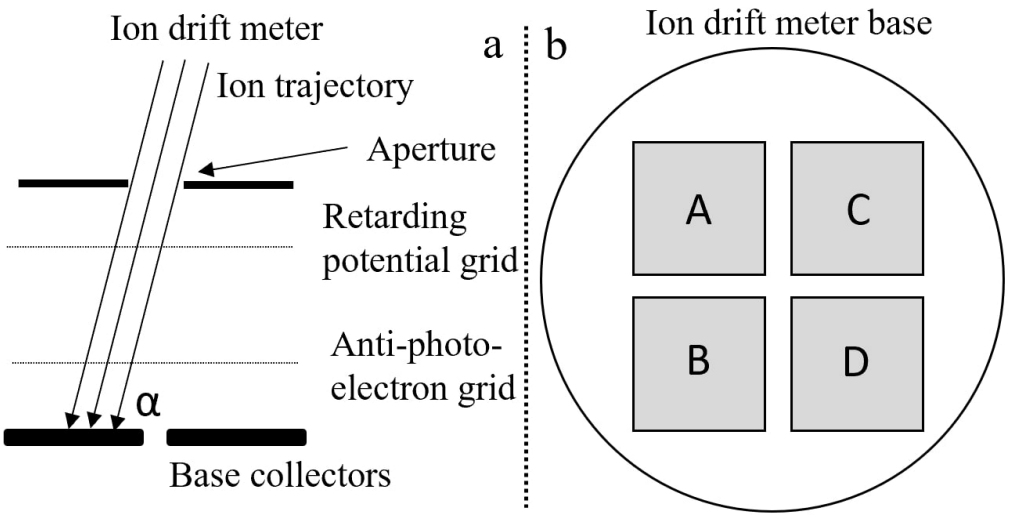

Figure 4.1: Illustration of an ion drift meter with integrated retarding potential analyzer. The side view in panel (a) illustrates a cross section of the aperture, grids, and collector plates. Panel (b) illustrates the four collectors at the base of the sensor.

understanding of our near-space environment, which in turn can lead to improved mitigation measures in case of extreme events. Ionospheric winds can also provide information on solid Earth activity such as earthquakes, volcanic eruptions, or high yield underground explosions [53–61]. Two types of waves are being considered in relation to earthquakes. Post seismic acoustic and gravity waves have been observed with satellites in low Earth orbit, and their connection with solid Earth phenomena is well understood from theory and computer simulations [53–55, 58]. Direct observations and statistical analyses have also been reported to support the occurrence of electromagnetic wave signatures prior to large earthquakes [56, 57, 59–61]. While not yet demonstrated, the possibility of observing ionospheric perturbations prior to large earthquakes remains a topic of vital interest, especially in countries located in seismically active parts of the planet (ibid).

Several designs of plasma flow meters have been used on satellites to measure ionospheric winds, including retarding potential analyzers, ion drift meters, "top hat" analyzers, ion imagers, and segmented Langmuir probes. The first satellites equipped with ion drift meters were deployed in the 1960s and 1970s [62–67]. Drift meters

are also used in several recent missions [68–72]. While the names differ, the working principles are similar. A simplified schematic of such a device is shown in Fig. 4.1. In this configuration, ions enter the sensor from the top aperture, and are collected by four current collectors at the base. The ram speed is measured with a retarding potential analyzer from which ion mass compositions, densities, temperatures and satellite potentials can also be determined [64, 70, 73, 74]. As shown in panel (a) of the figure, the voltage applied to the top grid is swept so as to block ions with varying energies from entering the sensor. As voltage is increased, abrupt drops are measured in the collected currents [73]. The voltages at which these reductions occur, correspond to different energy to charge ratios of incoming ions, in the satellite reference frame. The magnitude and shape of these drops also provide information on ion temperatures and relative densities. The second grid is biased to a fixed negative voltage to prevent the escape of photoelectrons from the base collectors. When collectors are exposed to UV radiation, simulations suggest that most photoelectrons are reflected back to the collector from which they were emitted [71]. The angle of incidence α of the plasma flow is determined from the relative currents collected by the segments. This, combined with the ram speed measured with the retarding potential analyzer, is used to determine the transverse flow velocity. The retarding potential analyzer/ion drift meter is robust, and it was used in many space missions. For example, VEIS on the WIND spacecraft was used to study the foreshock subsonic particles reflected from the bow shock [68]. This instrument can also be used to measure electron energies by reversing the analyzer electric field polarization. Similarly, IAP on DEMETER was used to measure plasma flow velocities with particular attention to the perturbed flow induced by waves caused by seismic activity [70]. The accuracy of ram speed measurements, obtained with IAP on DEMETER, was estimated to be approximately 10%, based on laboratory calibrations and computer simulations [70, 75]. Similar flow meters are also used on spacecraft, such as Dynamics Explorer B [67], C/NOFS satellite [71], and Ionospheric Connections Explorer [72]. A similar

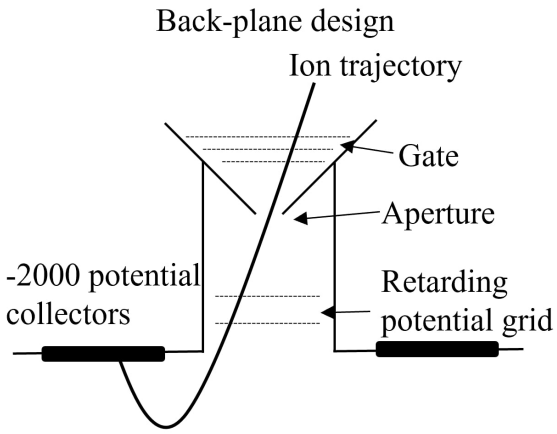

Figure 4.2: DIDM on the CHAMP satellite uses a back-plane design of ion drift meter. Ions are deflected 180° once they are in the detector dome using a -2000 volts potential.

instrument has been developed based on the same basic principle, referred to as the "backplane design". In this configuration, ions travel to the base of the sensor and are deflected by a strong electric field, to be collected on the backside of collector segments, as illustrated in Fig. 4.2. This configuration was used in DIDM on the CHAMP satellite to prevent direct UV radiation from entering the collectors, and minimize perturbations from photoelectrons [69].

The "top-hat" analyzer shown in panel a of Fig. 4.3 is widely used to sample charged particles over 360° in azimuthal [76–78]. Trajectories of incoming particles are bent by a radial electric field between two hemispherical electrodes of different radii. For a given potential difference between the two hemispheres, only particles in a narrow range of energy to charge ratio can follow a trajectory leading to the base collectors. The energy spectrum of the particles is then obtained by sweeping the potential difference between the two analyzer hemispheres. The "top hat" analyzer provides a pitch-angle range over the full 2-dimensional plane through the analyzer aperture. Ion imagers are yet another type of flow meter in which, as illustrated in panel b of Fig. 4.3. Ions enter through an aperture, and are dispersed by an

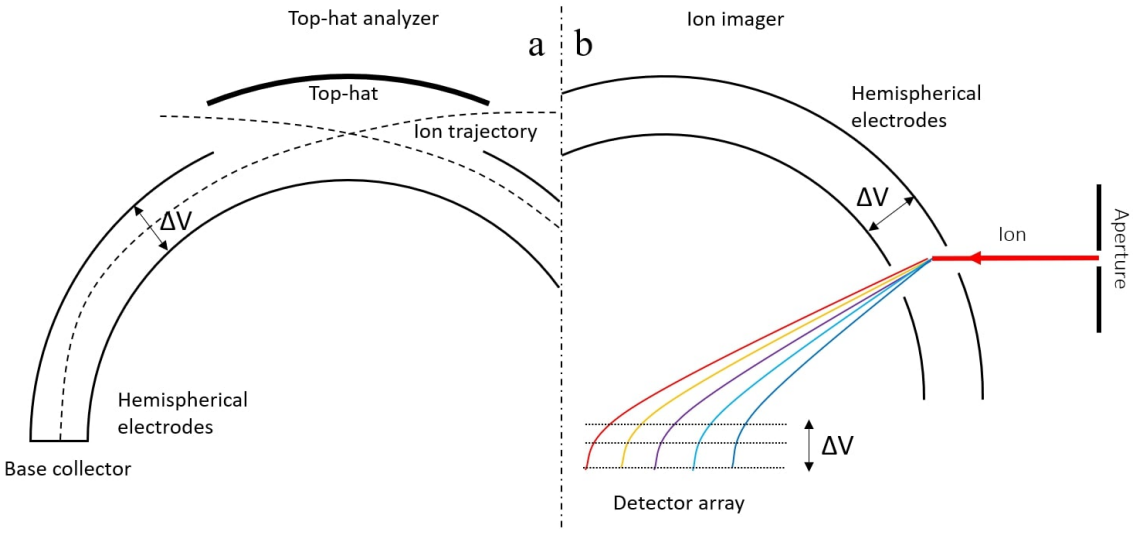

Figure 4.3: Illustration of a 'top hat' analyzer in panel (a) and ion imager in panel (b). Both devices use electric field to deflect incoming ion trajectories, and can sample ions over 360° azimuthal angles.

electric field between two concentric hemispherical shields, onto a detector array, as determined by their energy to charge ratio [79–83]. Depending on the setup, incoming particle velocities are measured over 180°, or the full 360° degrees in azimuthal angles. For example, the F3C Cold Plasma Analyzer (CPA) instrument on Freja can sample ions over a range of 360° in azimuthal angles around the satellite [79]. On Swarm, the Electric Field Instrument (EFI) consists of two imagers, each with 180° wide apertures, oriented perpendicularly to one another, thus providing a three-dimensional sample of incoming ion distributions [83]. Other spacecraft are also equipped with ion imagers, including ePOP [82], and Plante-B [80]. In principle, ion imagers can accurately measure ion drift velocities and ion masses without the need for sweeping voltage. In space, the Swarm ion imager performance vary due to several factors, including satellite potentials, changing plasma conditions, and attitude control [83]. The accuracy is estimated to be around 200 m/s, which should increase to about 100 m/s as more experience is gained using the instrument in orbit (ibid).

Segmented Spherical Langmuir Probes have also been used to measure bulk plasma

flow [84–86]. The surface of the probe is divided into several equipotential spherical caps or segments facing different directions, from which individual currents are measured. The relative currents from these segments and the supporting sphere can in principle be used to infer plasma density, temperature, ion masses and plasma flow velocity [84, 85]. This instrument was used on satellites such as DEMETER and Proba-2 [84–86]. It is also possible to infer plasma flow velocities indirectly from measured electric fields and the relation for the $\vec{E} \times \vec{B}$ drift. Boom-supported electric field probes are used on numerous satellite and rocket experiments, including ICE on DEMETER [87], Fields Instrument on FAST [88] and EFW on Cluster [89]. At lower latitudes, it is also possible to measure the neural wind speed in the ram direction, from the Doppler shift in atmospheric emission lines using an interferometer with laser beams [90].

One important difference between flow meters and more familiar Langmuir probes is that several theories have been developed to interpret measurements made with the latter, while no theory exists for the former. As a result, the inference of plasma flow velocities from flow meters must rely on laboratory calibration and computer simulations. Thus the goals of this study are to i) characterize the response of a proposed simple flow meter applicable to ionospheric wind, using computer simulations, ii) construct inference models based on multivariate regression, and iii), assess their predictive skills for conditions representative of the lower ionosphere. In the remainder of this paper, we present the geometry of a plasma flow meter, which should combine simplicity, robustness, and accuracy. The performance of the proposed instrument is assessed based on a combination of synthetic data constructed with computer simulations, and multivariate regressions. The simulation techniques, the sensor geometry, and the regression approaches are presented in Section 2. Simulation results and assessments of inference skills are presented in Section 3. The final section summarizes our findings and contains some concluding remarks.

4.2 Methodology

The segmented flow (SF) meter geometry considered is shown in Fig. 4.4. It is sufficiently compact to be mounted on small satellites such as attitude stabilized CubeSats. In the satellite reference frame, ions are incident from the ram direction, with speeds approximately equal to the satellite orbital speed. Thus, the meter needs to be mounted on the ram face of the satellite to allow ions to enter the aperture. In the proposed design, there are a total of 19 collecting segments, from which individual currents are measured. The top ring aperture is biased to -4 V with respect to the spacecraft in order to repel electrons and attract ions into the cone. This negative voltage at the top also serves to increase the radial dispersion of entering ions. All other segments at the base are biased to $+3$ V in order to i) enhance dispersion of the ion beam penetrating the sensor, and ii) retain photoelectrons that might be emitted, should solar UV enter the cavity. Enhancing radial spread of the incident ion beam at the base should make the distribution of collected currents more sensitive to the ion mass distributions and hence, to the ion effective mass. The curved conical faces of the sensor, both inside and outside, are assumed to be grounded to the satellite bus, implying that they would also be at the satellite potential V_s with respect to background plasma. The following paragraphs describe the approaches used to characterize the response of the flow meter to diverse space environment conditions, to construct models to infer physical parameters of interest from measurements, and to assess their predictive skills.

4.2.1 Symmetry

One key feature of the device considered is symmetry. In order to characterize the response of the multiple sensors to flows with components transverse to the cone axis, we need to carry out many three-dimensional kinetic simulations assuming different plasma parameters, consisting of densities, temperatures, ion compositions, flow ve-

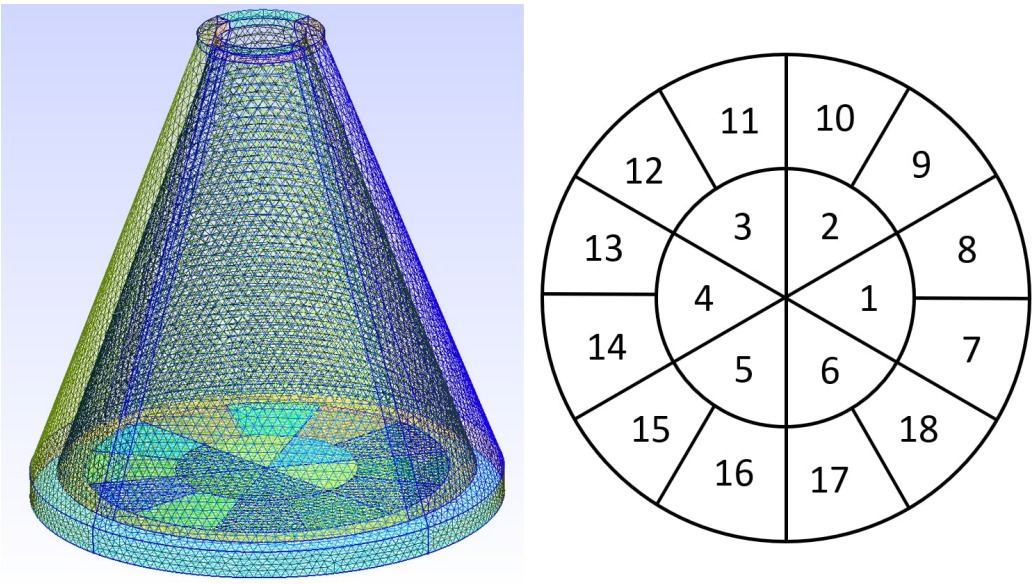

Figure 4.4: Illustration of the 3D geometry of the SF meter (left), and the 18 sectors at the base (right). The conical shell has a height of 5 cm, the outer radius at the base is 2.3 cm, and that at the top ring is 0.7 cm.

locities, and satellite potentials. These simulations are used to construct a solution library consisting of collected currents by each of the 19 segments, with corresponding space-plasma parameters. Without symmetry, simulations would be required for transverse flows covering the full 360° around the sensor axis. With the six-fold rotational symmetry, and the mirror symmetry in the 18 collecting segments at the base of the sensor seen in Fig. 4.4, however, simulations are only needed in a much smaller 30° angular sector. For example, simulations can be carried out to calculate currents collected by all segments, for flow velocities with transverse velocities in only the 30° sector 8. These currents can then be mirror imaged with respect to the horizontal axis between sectors 7 and 8 (or 13 and 14), to extend results to transverse velocities directed in sector 7. From there, the six-fold rotational symmetry can be used to further extend our simulation results to transverse velocities in all sectors, covering the full 360° of azimuthal angles; thus reducing the number of simulations by a factor 12 compared to what would be needed in the absence of symmetry.

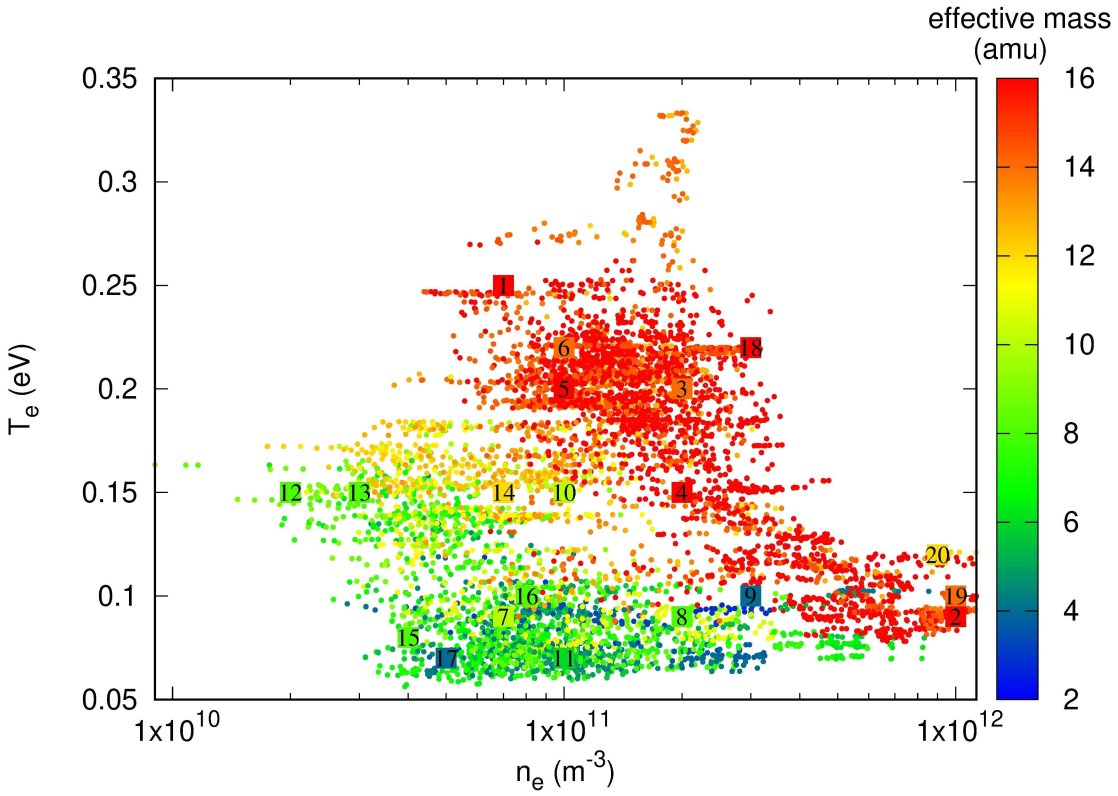

Figure 4.5: Scatter plot of plasma parameters obtained from the IRI model, corresponding to different latitudes, longitudes, altitudes, and times, as listed in Table 4.1. The x and y axes, and the color bar refer respectively to the electron density and temperature, and the ion effective mass. Numbered squares identify parameters used in the kinetic simulations.

4.2.2 Kinetic Simulations

The response of the sensor to different ionospheric wind conditions is simulated using the three-dimensional PIC code PTetra [19, 20]. In this model, space is discretized with unstructured adaptive tetrahedral meshes [21, 22], and Poisson's equation is solved at each time step, using Saad's GMRES sparse matrix solver [91]. Electrons and ions are treated kinetically, accounting for their physical masses, and particle trajectories are calculated self-consistently using computed electric fields. The parameters assumed in the simulations have been selected so as to be representative of ionospheric conditions encountered by satellites in low Earth orbit at mid, and low latitudes. A sample of electron and ion temperatures, electron densities, and ion mass distributions was obtained from the International Reference Ionosphere (IRI) [23] model for different latitudes, longitudes, altitudes, and times. The result is shown in Fig. 4.5, with points in the density-temperature scatter plot, and colors indicating ion effective masses. The numbered squares in the figures identify the twenty sets of plasma parameters ($T_e,\ T_i,\ n_e,\ m_{ieff}$) for which simulations were made. For each of the selected set of plasma parameters, several simulations were made for different satellite potentials, incoming plasma ram speeds, and transverse velocities distributed in the 30° sector 7, for a total of 310 simulations. The detailed plasma conditions and simulation results are given in [50]. When extended to the full 360° circle as described above, this produces a solution library consisting of 2676 entries used to train and assess our inference models.

For simplicity, and considering that in the conditions considered, O^+ and H^+ constituted 94% or more of all ion species, only these two ion species were considered in the simulations. Earth magnetic field is not accounted for, owing to the fact that typical ion gyroradii in the ionosphere are of order 1 m for H^+, and 4 m for O^+, which are much larger than the $\sim$ 5 cm size of the sensor considered. Secondary electron

Table 4.1: Range of ionospheric conditions considered with the IRI model, and corresponding ranges in space plasma parameters.

Environment and plasma conditions	Parameter range
Years	1998 2001 2004 2009
Date	Jan 4 Apr 4 Jul 4 Oct 4
Latitude	$-65°$ - $+65°$ with increment of $26°$
Longitude	$0°$ - $-360°$ with increment of $30°$
Hours	0-24 with increment of 8 hours
Altitude	450-550 km
Ion temperature	0.07-0.12 eV
Electron temperature	0.09-0.25 eV
Effective ion mass	4-16 amu
Density	$2 \times 10^{10} - 1 \times 10^{12} \mathrm{m}^{-3}$
Ram velocity	7000 -8000 m/s
Transverse speed	0-500 m/s
Angles	0-30°
Spacecraft potential	-2-1 eV

emission is ignored in the calculations because of the low electron temperatures (below 0.5 eV) encountered in the regions of interest. Photoelectron emission is also not taken into account, which is justified when the satellite is on the night side of its orbit or when the meter aperture is not exposed to solar illumination.

The Debye length in the background plasma can range from 2 mm to 3 cm for the plasma conditions considered. Due to the negative potential of the top ring, however, electrons are strongly repelled, and the electron density in the vicinity of the top aperture is much smaller than that of the background plasma. As a result, the region near the circular aperture is practically devoid of electrons. This, combined with the high speed of incoming ions in the satellite reference frame, leads to negligible Debye

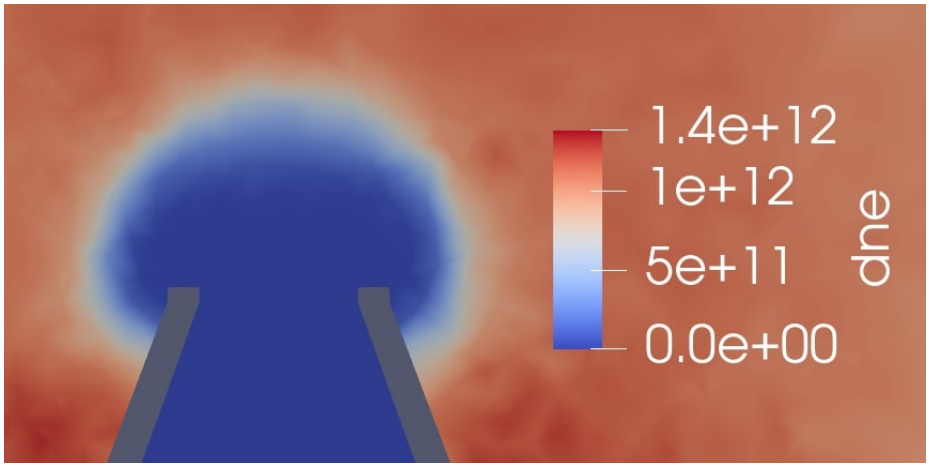

Figure 4.6: Example cross section of electron density around the top ring. This corresponds to plasma condition 19 in Fig. 4.5, with n_e, m_{ieff}, T_e, T_i being 1×10^{12} m^{-3}, 14 amu, 0.1 eV and 0.08 eV respectively, and a satellite potential of +1 V.

shielding in the immediate vicinity of the aperture, resulting in a potential across the surface delimited by the ring being no less in absolute value, than approximately 35 % of the the ring potential with respect the background plasma. This is illustrated with a cross section of the electron density profile near the top ring computed with plasma conditions 19 in Fig. 4.6, assuming a satellite potential of +1 V. In this case, the Debye length in the background plasma is approximately 2 mm, and yet, electrons are effectively repelled near the top ring, and shielding in the plane of the ring is clearly negligible. Simulations indicate however, that if the satellite potential V_s is larger than $\sim$ 1 V which could happen when the satellite is exposed to solar radiation or carrying active experiments [92–94], the base sensors start collecting a noticeable amount of electrons passing through the top aperture. This in turn would affect the measurements. In the lower ionosphere at mid and low latitudes where photoelectron and secondary electron emission are not significant, a spacecraft should be charged negatively [95]. In the following, the proposed sensor response is assessed assuming spacecraft potentials ranging from -2 to 1 V.

4.2.3 Multivariate regression

Given a solution library, the next step is to construct models capable of inferring plasma parameters from measurements. In the following, we describe two approaches for constructing such models, which will be applied and assessed for their inference skills in Sec. 4.3. Several approaches are possible, including empirical parametric fits and multivariate regressions [96]. Here we use two regression approaches based on i) Radial Basis Functions (RBF), and ii) Deep Learning Neural Networks. In either case, two steps are involved in the construction of a model. The first step consists of training a model on a subset of the solution library, the "training set", while the second step consists of applying the trained model to a distinct data set, the validation set consisting of the remaining subset of the library. Considering that the models are trained by optimizing their inference skill on a training set consisting of a subset of the our solution library, and that in validation, it is applied to a different subset without further optimization, it follows that predictive skills are generally better on the training set than on the validation set. Model skills applied to the training set can be improved by further refining the model, but improvements in training do not necessarily correspond to improvements in validation inferences. Beyond a certain level of refinement in training, "overfitting" occurs, and inference skill degrades for the validation set. A good model is one with the right level of refinement so as to provide the best inference skill when applied to the validation set. Let us now briefly present the two regression methods used in our study.

Radial Basis function

Radial basis Function is one of the most basic regression techniques, and it is applied in many fields, including image mapping, and data tracking [44]. Given a set of independent vectors $\vec{X}$ and corresponding dependent vectors $\vec{Y}$, a general expression

for RBF regression is given by

$$\vec{Y} = \sum_{i=1}^{N} a_i G\left(\|\vec{X} - \vec{X}_i\|\right), \tag{4.1}$$

where $\vec{Y}$ is a vector of parameters to be inferred, $\vec{X}$ is a vector consisting of independent, measured quantities, and $(\vec{X}_i, \vec{Y}_i)$ are reference nodes or pivots in the space of independent-dependent variables. G is a function of a real variable, a_i are fitting coefficients, and N is the number of pivots used in the regression. In RBF, $\vec{X}$ and $\vec{Y}$ can be vectors of different dimensions. In what follows, however, dependent variables $\vec{Y}$ will always be scalars (vectors of dimension one), and $\vec{X}$ will be vectors of different dimensions, depending on the physical parameter being inferred. In Eq. 4.1, the argument of G is the L^2 norm, or Euclidean distance between $\vec{X}$ and $\vec{X}_i$; whence the name "radial" in RBF. The choice of G is arbitrary, provided that, for a given set of pivots, the set of N interpolating functions in Eq. 4.1 be independent of one another. When constructing a regression model with RBF, the function G, and the number and distribution of pivots must be chosen so as to yield the best possible predictive skill for a given problem. Two G functions have been found to give good predictive skill for the inferred physical parameters considered. They are described with the physical parameters in Sec. 4.3. The number and distribution of pivots have similarly been selected so as to provide optimal accuracy when inferring dependent parameters in a data set. Two types of cost functions have been considered, the maximum absolute error (MAE):

$$\epsilon_{abs} = Max \mid Y_{sim} - Y_{mod} \mid, \tag{4.2}$$

and the maximum relative error (MRE):

$$\epsilon_{rel} = Max \left| \frac{Y_{sim} - Y_{mod}}{Y_{mod}} \right|, \tag{4.3}$$

calculated over a given data set, where Y_{sim} are known plasma parameters used in the simulation such as density, and Y_{mod} are the model-inferred parameters.

In order to carry out this task and construct a model, the fitting coefficients a_i in Eq. 4.1 have to be determined. This is done first by requiring collocation of inferred and known parameters at pivots; that is, by solving the set of equations

$$\sum_{j=1}^{N} a_j G(\|\vec{X}_i - \vec{X}_j\|) = \vec{Y}_{i,sim},\ i = 1, N. \tag{4.4}$$

Given a training data set of $\mathcal{N}$ nodes, the selection of N pivots is made by constructing models for all possible $\mathcal{N}$ choose N combinations of pivots among the $\mathcal{N}$ nodes, and selecting the one which minimizes the cost function. When the best distribution of pivots is found, the model can be further improved by relaxing collocation, by allowing for small deviations from the $\vec{Y}_{i,sim}$ and minimizing the cost functions with respect to these deviations. Yet another improvement is to go over all $\mathcal{N}$ choose N possible combinations of pivots in parallel on n processors, in such a way that each processor goes through different combinations. In this case, each processor finds its unique best combination of pivots. One obvious advantage of this is an increase in speed. Another one is that relaxation, or accounting for the "nugget effect", can be applied to each of the distinct n best combinations, and selecting the combination which, after relaxation, produces the smallest cost function. It is found that the best combination then, is not necessarily the one that minimizes the cost function before relaxation. With this strategy, and using several processors, it is possible to reduce the cost function in a training set by several %, compared to a minimization made without relaxation.

Given the size of the data, $\mathcal{N}$ choose N can be very large. One strategy is to combine RBF with the Monte Carlo method to do a non-exhaustive search for the model. In this approach, a small subset (e.g. 100 entries) is picked each time randomly from the training data set to train a model, then the model is applied to the entire training data set to calculate the cost function. The best model is selected after a certain time and it is applied to the validation data set to determine the validation

error.

Neural network

Neural networks have increasingly been proven useful in many applications, including plasma physics and space physics [97, 98]. In this work, we use feedforward deep learning networks to infer plasma parameters from currents collected from the 19 segments in the proposed flow meter. An illustration of a feedforward network is shown in Fig. 4.7, with the input layer, hidden layers, and the output layer. In our problem, each node in the input layer is assigned a current from one of the segments. Node j in layer i is assigned a value $u_{i,j}$, and each node of the next layer $i+1$ is "fed" by all the nodes of the previous layer according to

$$u_{i+1,k} = f\left(\sum_{j=1}^{n_i} w_{i,j,k} u_{i,j} + b_{i,k}\right), \tag{4.5}$$

where $w_{i,j,k}$ are weight factors, $b_{i,j}$ are bias terms, and f is a nonlinear activation function. In this study, the bias terms are all set to zero. The w coefficients are first generated using the Monte Carlo method, and then gradient descent is used to further decrease the cost function over the training data. Training sets consisting of 500 data entries are used to train neural network models. As with RBF, many models are trained before the final model is selected. The models are then applied to the validation data sets to obtain the validation error.

4.2.4 Noise

Given a trained model, the skill and robustness of inference are tested against noise in the validation sets. Noise in collected currents can be statistical in nature, or it can be associated with physical processes such as waves and turbulence. The current collected by a segment is given by the number of particles N collected in a given sampling time τ, multiplied by their respective charges, and divided by τ; that is,

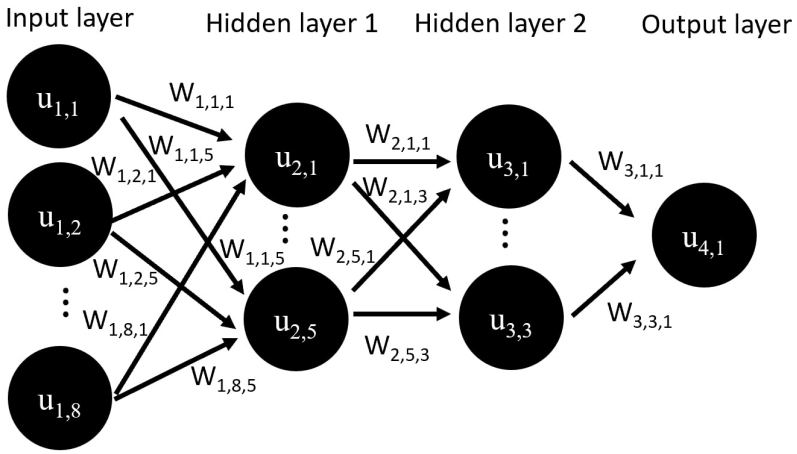

Figure 4.7: Schematic of a feedforward neural network.

assuming singly ionized ions for simplicity,

$$I = \frac{Ne}{\tau}. \tag{4.6}$$

Owing to the discrete nature of this process, the number N follows approximately Poisson statistics. The standard deviation; that is, the noise level, in N is therefore approximately the square root of $\bar{N}$, the average value of N: $\sigma_N \simeq \sqrt{\bar{N}}$. Thus, it follows that the standard deviation in the collected current is approximately

$$\sigma_I \simeq \frac{\sigma_N e}{\tau} \simeq \sqrt{\frac{Ie}{\tau}}. \tag{4.7}$$

In simulations however, the number of simulation particles N_s accounted for, is generally smaller than the actual number of physical particles in a plasma. In order to account for that, simulation particles carry a statistical weight w, corresponding to the number of actual particles that they "represent". Currents calculated in simulations are therefore obtained by multiplying the charge of each collected particle by its statistical weight as in

$$I = \frac{wN_s e}{\tau}, \tag{4.8}$$

and the resulting standard deviation in the current calculated in a simulation is

$$\sigma_I \simeq \frac{w\sigma_N e}{\tau} \simeq \sqrt{\frac{wIe}{\tau}}. \tag{4.9}$$

The standard deviation in the collected current can also be calculated directly from our simulation results, by considering a case with zero transverse flow velocity. In this case, by symmetry, all six inner segments should collect the same current, as should the twelve outer segments. Thus, calculating the standard deviation in these currents provides an estimate of the intrinsic statistical noise in the current collected by a single segment. For example, in one of the simulations, using a sampling time of 1μs, in which ions have a statistical weight $w = 2$, the average current per inner segment is calculated to be $I \simeq 2n$A. In this case, the standard deviation of the current over the six segments is found to be $\simeq 29$ pA, which is in good agreement with the 25 pA estimated from Eq. 4.9.

In order to test the robustness of our models, additional noise is introduced in our validation sets, in addition to the intrinsic statistical noise mentioned above. Here again, this added noise is assumed to be proportional to the square root of the collected current as per

$$I_\sigma = I_0 \left(1 + r\sigma \sqrt{\frac{I_0}{1n\text{A}}} \right), \tag{4.10}$$

where I_σ is the current collected with added noise, I_0 is the simulated collected current from the solution library for a given segment, σ is a relative standard deviation, and r is a zero-mean random number with Gaussian distribution and unit standard deviation. For each value of σ, 100 sets of random noise have been used to calculate the averages of the maximum errors and Root-Mean-Squared (RMS) errors reported in Tables 4.3.

4.3 Results and discussion

We now proceed with the construction of models for selected plasma parameters.

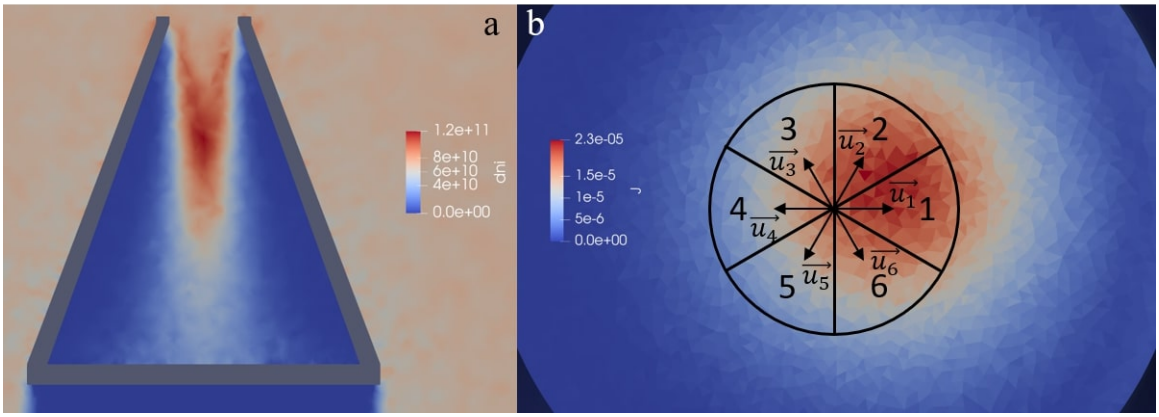

Figure 4.8: An example of cross section of the ion density in and out of the SF meter (a), and collected current density profile at the base (b). The density is in units of m^{-3}, and current density in units of Am^{-2}. This corresponds to condition 14 in Fig. 4.5, with n_e, m_{ieff}, T_e, T_i, satellite potential, ram and transverse plasma flow speeds being 7 $\times 10^{10}$ m^{-3}, 12 amu, 0.15 eV, 0.11 eV, 0 V, 7500 m/s and 500 m/s respectively. Results from simulations made with different parameters shows similar qualitative features.

4.3.1 Transverse flow velocity

The inference of transverse velocities relies on the symmetry and the currents collected by the base 18 segments as described Sec. 4.2.1. This is made in two steps in which i) the direction of the transverse flow velocity, and ii) its magnitude are determined.

Transverse flow direction - The vector approach

An obvious manifestation of a transverse flow velocity in incident plasma is an azimuthal asymmetry in the currents collected at the base of the sensor, as shown in Fig. 4.8. Given the geometry of the sensor, the shift in the centroid of the collected current must be in the direction of the transverse plasma flow velocity. This shift in turn can be determined from the average of the unit vector pointing in the middle of each sector, as shown in panel b of Fig. 4.8, weighted with the current that it collects. In practice, two averages are made, for the inner sectors as

$$\vec{U}_1 = \sum_{i=1}^{6} I_i \hat{u}_i, \tag{4.11}$$

Table 4.2: Examples of transverse wind angles obtained from $\vec{U}$ in the vector approach. Each run number corresponding to a set of plasma conditions mentioned in section 4.2.2. The satellite potential in both cases is zero. "Inner", "Outer", and "Simulation" refer respectively to the inner ring vector, the outer ring vector and the wind direction used in the simulation.

Plasma condition#	Wind speed (m/s)	Inner	Outer	Simulation
1	125	18.8°	17.6°	10°
1	250	12.2°	13.0°	10°
1	375	12.4°	12.2°	10°
1	500	10.5°	11.9°	10°
2	125	28.4°	30.8°	20°
2	250	23.7°	23.6°	20°
2	375	23.3°	23.0°	20°
2	500	21.0°	22.8°	20°

and a similar expression is used for $\vec{U}_2$, calculated with the 12 outer sectors. The directions of the two vectors give indications of directions of the wind, as shown in Table 4.2. These vectors are then combined linearly as:

$$\vec{U} = (1 - \alpha)\hat{U}_1 + \alpha\hat{U}_2, \tag{4.12}$$

where the parameter α is selected so as to minimize the absolute error in the inferred transverse velocity over a given training data set. $\alpha \simeq 0.94$ is found to be optimal in all cases considered, and it is the value used in the inference models considered below.

Transverse flow speed and velocity

Given a direction of the flow from Eq. 4.12, the transverse velocity can be obtained from the transverse speed. The magnitudes of $\vec{U}_1$ and $\vec{U}_2$ have a strong dependence on the density of the plasma. In practice, it is found that if $\vec{U}_1$ is normalized by the sum of inner base currents, and $\vec{U}_2$ by the sum of total base currents, the dependence

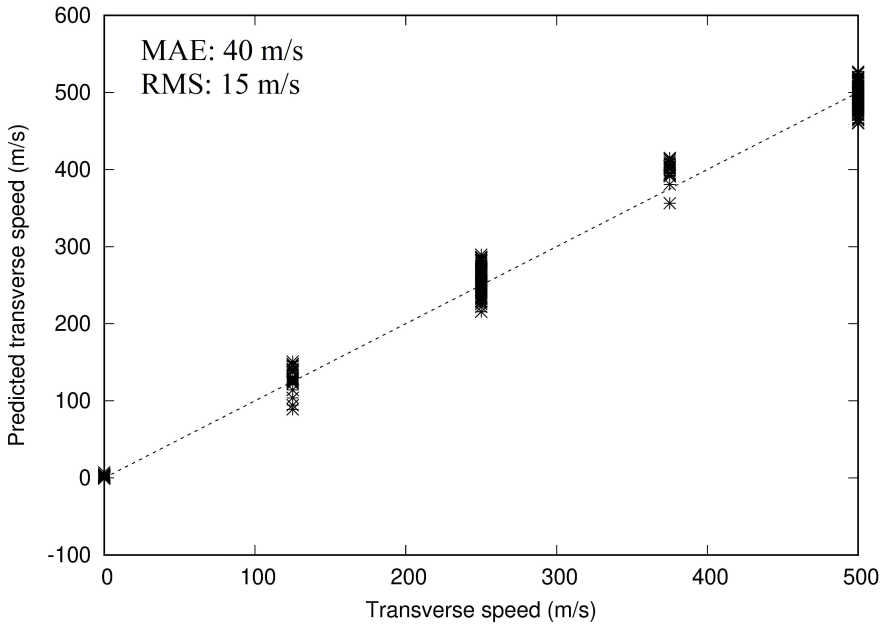

Figure 4.9: Correlation plot of the transverse wind speeds inferred for the validation set, vs. actual speeds used in the simulations. For reference, the dotted line corresponds to a perfect correlation. In this case, RBF is used with 5 pivots, leading to a maximum absolute error (MAE) of 40 m/s, and a RMS error of 15 m/s.

on the density is greatly reduced, and more accurate values of the speed can be inferred. We then use the normalized magnitudes of $\vec{U}_1$ and $\vec{U}_2$ as the two components of independent vectors $\vec{X}$ in RBF to infer transverse flow speed. For example, a correlation plot of inferred speeds as a function of the actual speed from the solution library is shown in Fig. 4.9. In this case, the model is constructed with a training set of 1338 randomly selected nodes from the solution library, using five pivots as explained in Sec. 4.2.3, and it is applied to a validation set consisting of the 1338 remaining nodes. The regression function used here is $G(x) = 0.5x^{1.6} \times log(x^2)$ for $x > 0$ and the cost function is the maximum absolute error over the set considered. The figure also shows the value of the cost function (40 m/s) and the RMS error (15 m/s) computed on the validation set. Figure 4.10 shows predicted transverse velocities and actual transverse flow velocities without (a) and with (b) 2% ($\sigma = 0.02$) added statistical noise in the validation set using Eq. 4.10. Here the model uses the same training

Table 4.3: Errors in inferred angles, transverse speeds, velocities, and densities calculated without, and with noise added to currents in the validation set.

Parameter:	Angle (°)	Speed (m/s)	Velocity (m/s)	Density (%)
Method:	Vector	RBF	Vector+RBF	RBF
Skill metric:	RMS	RMS	RMS	RMSrE
$\sigma = 0$	3.2	15	20	11
$\sigma = 1\%$	3.6	16	20	12
$\sigma = 2\%$	4.2	17	21	12
Skill metric:	MAE	MAE	MAE	MRE
$\sigma = 0$	10.7	40	45	23
$\sigma = 1\%$	15	52	58	32
$\sigma = 2\%$	20	70	75	49

and validation sets as for Fig. 4.9. When the model is applied to the validation set, the maximum absolute error, and root-mean-squared error are 45 m/s and 20 m/s respectively, when no noise is added. These errors increase respectively to 75 m/s, and 21 m/s when 2% relative noise is added to the validation set, which corresponds to approximately 80% of the simulation statistical noise estimated from Eq. 4.9. Results from neural network, not shown here, are comparable within 30%, with RBF prediction being slightly more accurate. More inference skill metrics are listed in Table 4.3, for different levels of added noise. As expected, our model predictive skill decreases as noise is added, and the maximum absolute error is found to increase by a factor two for a level of added noise of approximately 2%.

4.3.2 Density Prediction

While our primary objective is to infer ionospheric plasma flow velocities, it is interesting to explore the possibility for the proposed instrument to be used to infer other physical quantities. This is motivated by the fact that the currents collected by the many segments in the meter, and their relative values, are sensitive to sev-

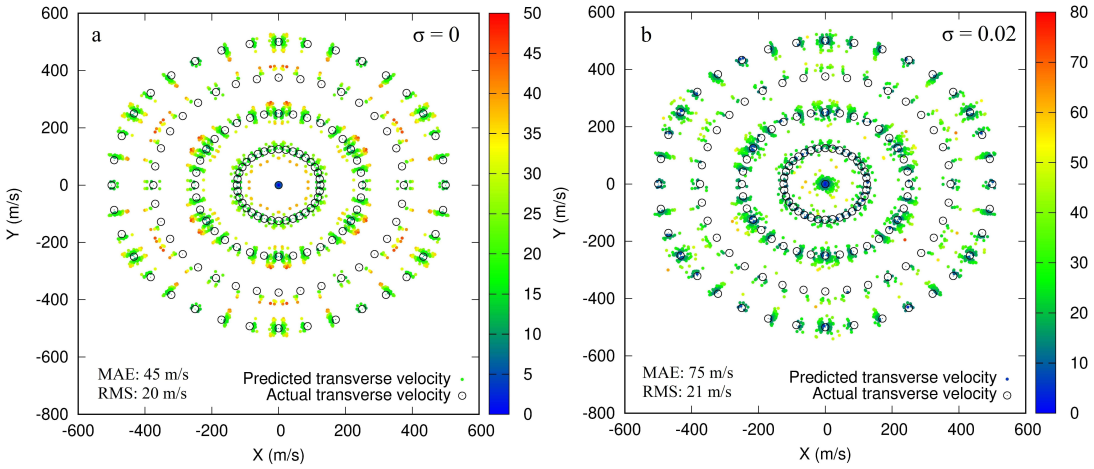

Figure 4.10: Actual and inferred transverse velocities without (a) and with (b) 2% added noise in the validation data set. The color scale shows the absolute errors in the model velocity predictions. Inferred velocities were obtained with RBF regression, using 5 pivots.

eral satellite plasma environment parameters, including ion densities and masses, ion temperatures, ram, and transverse velocities, and satellite potentials. Models were constructed for the plasma density using both RBF and neural network regression, and both are found to yield inferences with comparable skills. Here, however, considering the nearly two orders of magnitude range over which densities vary in our solution library and the fact that the density is a positive definite quantity, the cost function chosen in the construction of the models consists of the maximum relative error (in absolute value). This is preferred to the absolute error because, with the latter, models can be constructed with excellent skills for the larger densities, but poor ones for lower densities. Among the several G functions tested, the best one for predicting density was $g(\mathbf{x}) = \mathbf{x}^5$. Here, 5 pivots were used as a good balance between training and validation inference skills. 500 entries were used to train models using neural networks, with a four-layer network with 19, 15, 7, and 1 nodes. Figure 4.11 shows correlation plots of inferred density, as a function of actual densities obtained with neural network (a) and RBF (b) regression, for the validation set without the addition of statistical noise. Both regression techniques yield comparable predictive

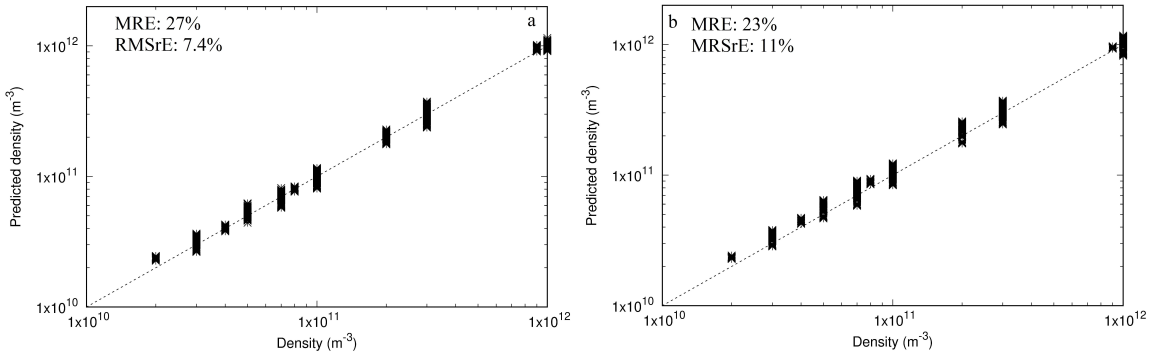

Figure 4.11: Predicted densities vs. densities used in simulations obtained by minimizing the maximum relative error. The neural network prediction with 500 points is shown on the left (relative error 27%) and the RBF predicted density using 5 pivots is shown on the right (relative error 23%). The dotted line corresponds to a perfect correlation between predictions and actual densities.

skills, with maximum relative errors of 27% and 23%, and root-mean-square relative errors of 7.4% and 11% respectively for the neural network and RBF. As for the transverse flow velocity, the models' robustness to statistical noise was assessed by adding random noise to the currents collected by each segment, as per Eq. 4.10. The impact on predictive skills is given in Table 4.3, which again shows a degradation of skill with an increase in the level of noise.

4.4 Summary and conclusion

Results are presented for a particle sensor, which could be mounted on satellites, to infer in situ transverse plasma flow velocities. The device consists of several electrically biased segments at the base of a conical enclosure, and a circular ring on the top aperture, from which ion currents are measured. Three-dimensional kinetic particle in cell (PIC) simulations are made to construct a solution library and data sets, for plasma environment conditions of relevance to satellites in low Earth orbit. The symmetry of the device enables the construction of data sets for transverse velocities directed in the full 360° in the plane perpendicular to the ram direction of plasma flow velocities, from simulations made in only a 30° sector. Owing to the

large computational resources needed to carry out kinetic simulations, symmetry is key in reducing the required number of simulations. Training and validation data sets, constructed with our solution library, are used to construct regression models capable of inferring transverse velocities and plasma densities. Two approaches are assessed for constructing such models, consisting of radial basis function, and neural network regressions. The two approaches are found to have comparable skills for inferring both transverse velocities, and plasma densities. With the configuration considered, it was not possible to make an accurate inference of the plasma flow speed in the ram direction. This is because the ion current collected by the base sensors is approximately proportional to the product of the density and ram speed, and the variation in ram speed is small compare to the orbital speed. Better inference of the ram speed should nonetheless be achievable by using a separate, or integrated retarding potential analyzer as illustrated in Fig. 4.1.

The level of statistical noise in the collected currents, associated with the discrete nature of kinetic simulations, explains in part the relatively small discrepancies between our model predictions and actual values in the data sets. Considering that simulations are made with significantly fewer particles than there would be in an actual plasma, the statistical uncertainties in our simulated currents are larger than those that would occur in space under similar conditions. The tolerance of our models to statistical noise is assessed by adding varying levels of normally distributed noise to the currents in our validation sets, in addition to the numerical simulation noise mentioned above. The skill of both RBF and neural network regressions decreases as noise is added, and it is estimated that an additional 2% relative noise leads only to approximately doubling in the uncertainty of model inferences in both cases.

Several approximations were made in the simulations used to construct our training and validation sets. In particular, the presence of a satellite bus was not taken into

account, which is justified if the flow meter is mounted on the ram face of a satellite, and the fact that satellites in low Earth orbit have supersonic ram velocities. The geomagnetic field was also neglected, which is justified by the fact that typical ion thermal ion gyro-radii is a factor 10 or more, larger than the size of the sensor. The neglect of solar illumination and photoelectron emission is valid when the satellite is on the night side of its orbit or when no solar radiation enters the sensor. When the satellite is sunlit, however, it would be possible for the negatively biased ring at the sensor aperture, to emit photoelectrons which, owing to the negative bias, would be repelled, and appear as collected positive current. Solar UVs could also enter the aperture and reach directly, or indirectly through multiple reflections, the positively biased segments. This in turn would result in photoelectrons being emitted inside the flow meter which, owing to the positive bias of the segments at the base, would likely be attracted back to the segments, albeit, not necessarily at the exact position where they were emitted. This, and the exposition of the negative ring at the aperture, would likely affect measured currents, and require corrections in the models presented above to infer plasma parameters. These effects should be included in models constructed to support missions, in which specific spacecraft geometry, orbital parameters, and expected range of plasma environment parameters would be taken into account. Such an analysis is of course well beyond the scope of this preliminary study, as it would require accounting for a broader range of parameters and environmental conditions, and would require significantly more simulations. Considering the costs, efforts, and years of preparation preceding a space mission, such an investment, enabling better data acquisition, should nonetheless be well justified.

Chapter 5

m-NLP inference models using simulation and regression techniques

This chapter is based on a paper submitted to the Journal of geophysical research: space physics, entitled "m-NLP inference models using simulation and regression techniques", by Guangdong Liu, Sigvald Marholm, Anders J. Eklund, Lasse Clausen and Richard Marchand. This work was supported by the China Scholarship Council, the Natural Sciences and Engineering Research Council of Canada, the Research Council of Norway (Grant Agreement No. 275655 and 325074), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 866357, POLAR-4DSpace). The kinetic simulations used in this study were made on the Compute Canada computing infrastructure. S.M. also acknowledges Dag Mortensen, Øyvind Jensen, and the Institute for Energy Technology for permission to participate in this research.

5.1 Introduction

Langmuir probes are widely used to characterize space plasma and laboratory plasma. A variety of Langmuir probe geometries are being used, such as spherical [99], cylindrical [6], and planar probes [100–102]. Probes can be operated in sweep mode [85], harmonic mode [103], or fixed biased mode [3], for different types of missions and mea-

surements. Despite operational differences, all Langmuir probes consist of conductors exposed to plasma to collect current as a function of bias voltage. A common approach to infer plasma parameters from Langmuir probes is to sweep the bias voltage and produce a current-voltage characteristic, which can be analyzed using theories such as the Orbital Motion-Limited (OML) [8] theory, the Allen-Boyd-Reynolds (ABR) theory [104–106], and the Bernstein-Rabinowitz-Laframboise (BRL) theory [107, 108] to obtain plasma parameters such as density, temperature, and satellite floating potential. The temporal and, on a satellite, the spatial resolution of Langmuir probe measurements are determined by the sweep time, which varies based on the mission's scientific need and available resources. Considering the orbital speed to be around 7500 m/s for a satellite in low Earth orbit (LEO), the spatial resolution of sweep bias Langmuir probe can vary from tens of meters, to kilometers, depending on the sweep frequency. In order to study the formation of density irregularities that scale from meters to tens of kilometers at high and low latitudes, a sampling frequency of near 1 kHz is required [3, 4]. A solution, proposed by Jacobsen is to use multiple fixed biased needle probes (m-NLPs) to sample plasma simultaneously at different bias potentials in the electron saturation region [3]. This approach would eliminate the need for sweeping the bias voltage, and greatly increase the sampling rate of the instrument.

The first inference models for m-NLPs relied on the OML approximation, from which the current I_e collected by a needle probe in the electron saturation region is written as:

$$I_e = -n_e e A \frac{2}{\sqrt{\pi}} \sqrt{\frac{kT_e}{2\pi m_e}} \left(1 + \frac{e(V_f + V_b)}{kT_e}\right)^{\beta}, \tag{5.1}$$

where n_e is the electron density, A is the probe surface area, e is the elementary charge, k is Boltzmann's constant, T_e is the electron temperature, V_f is the satellite floating potential, V_b is the bias potential of the probe with respect to the satellite, and β is a parameter related to probe geometry, density, and temperature [4, 15]. Several assumptions were made in the derivation of this inference equation; such as

the probe length must be much larger than the Debye length, and the plasma is non-drifting. If these assumptions are valid, then $\beta = 0.5$, and as first suggested by Jacobsen, a set of m-NLPs can be used to infer the electron density independently of the temperature [3]. For a satellite in near-Earth orbit at altitudes ranging from 550 km to 650 km, we can expect a Debye length of around 2-50 mm, and an orbital speed of around 7500 m/s. A common length for m-NLP instrument used on small satellites is $\sim$ 25 mm [2, 6, 109], which is often comparable to, and sometimes smaller than the Debye length. In lower Earth orbit, ion thermal speeds are usually less than the orbital speed, while electron thermal speeds are usually higher than the orbital speed. Thus, the orbital speed is expected to mainly affect ion saturation region currents for Langmuir probes. However, electrons can only penetrate the ion rarefied wake region behind the probe as much as ambipolar diffusion permits [10]. As a result, electron saturation currents are also influenced by an orbital speed. One consequence is that the $\beta = 0.5$ assumption does not hold in Eq. 5.1, and a better approximation for the current is obtained with β values between 0.5 and 1. For example, in a hot filament-generated plasma experiment, Sudit and Woods showed that β can reach 0.75 for a ratio between the probe length and the Debye length in the range of 1 to 3. For larger Debye lengths, they also observed an expansion of the probe sheath from a cylindrical shape into a spherical shape [11]. Ergun and co-workers showed that with a ram speed of 4300 m/s in their simulations, the current collected by a 40.8 cm needle probe is better approximated with Eq. 5.1 using a β value of 0.67 instead of 0.55 calculated in a stationary plasma [16]. In the ICI-2 sounding rocket experiment, β calculated from three 25 mm m-NLPs varied between 0.3 to 0.7 at altitudes ranging from 150 to 300 km [4]. Simulation results by Marholm et al. showed that even a 50 mm probe at rest can be characterized by a $\beta \sim 0.8$ [14], in disagreement with the OML theory. In practice, needle probes are mounted on electrically isolated and equipotential guards in order to attenuate end effects on the side to which it is attached. The distribution of the current collected per unit length is nonetheless not

uniform along the probe, as more current is collected near the end opposite to the guard. A study by Marholm & Marchand showed that for a cylindrical probe length that is 10 times the Debye length, β is approximately 0.72. For a probe length that is 30 times the Debye length, β is approximately 0.62, and with a guard, this number is reduced to 0.58 [15]. Although this number approaches 0.5, 30 times the Debye length is a stringent requirement for OML to be valid, and it is hardly ever fulfilled in practice. Experimentally, Hoskinson and Hershkowitz showed that even with a probe length 50 times the Debye length, β is approximately 0.6, and the density inference based on an ideal $\beta = 0.5$ is 25 % too high [12]. Barjatya estimated that even a 10% error in β (to 0.55) can result in a 30 % or more relative error in the calculated density based on the $\beta = 0.5$ assumption [13]. In what follows, we find that densities estimated using Eq. 5.1 assuming $\beta = 0.5$ are about three times larger than the known values used as input in our simulations, as illustrated in section 5.3.1. This is consistent with findings in [13, 31], considering β calculated in our simulation is in the range of 0.75 to 1. Another approach proposed to account for the fact that β is generally different from 0.5, consists of determining the n_e, V_b, T_e and β, as adjustable parameters in nonlinear fits of measured currents as a function of voltages. This lead to remarkable agreement with density measured using a radio frequency impedance probe on the international space station [10, 110, 111]. This method was originally applied to a probe operated in sweep voltage mode, but it can be straightforwardly adapted to fixed bias m-NLP measurements [4, 10, 13].

In the following, we assess different techniques to infer plasma densities, and satellite potentials from fixed bias needle probe measurements based on synthetic data obtained from kinetic simulations. We also present a new method to interpret m-NLP measurements based on multivariate regression. Our kinetic simulation approach and the construction of a synthetic data set are presented in Sec. 5.2. In Sec. 5.3, regression models are trained using synthetic data sets, and they are assessed using distinct validation sets. In Sec. 5.4, the same models are applied to NorSat-1 data,

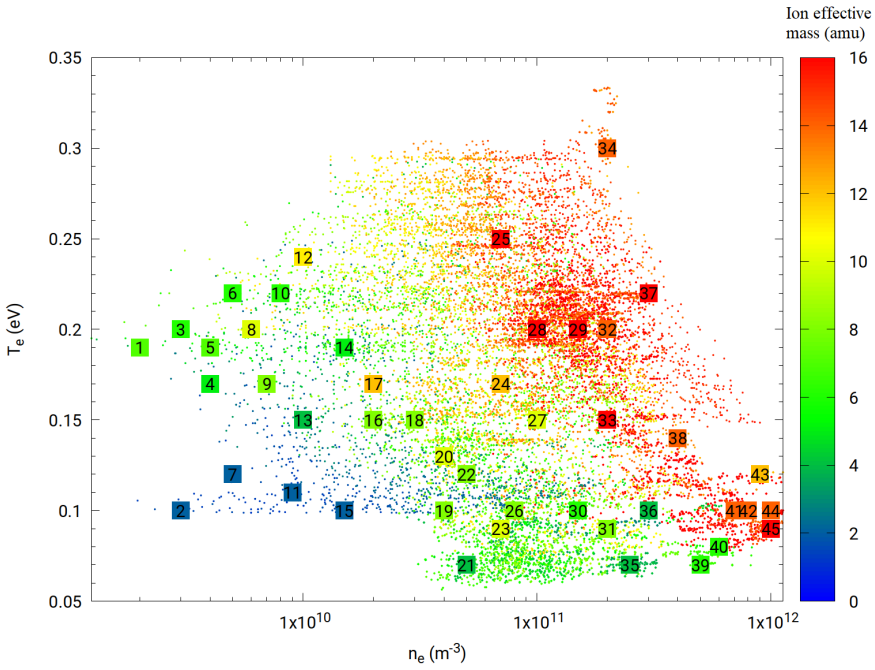

Figure 5.1: Scatter plot of plasma parameters obtained from the IRI model, corresponding to different latitudes, longitudes, altitudes, and times, as listed in Table 5.1. The x and y axes, and the color bar refer respectively, to the electron density, electron temperature, and the ion effective mass. Numbered squares identify the set of parameters used in the kinetic simulations.

to infer densities and satellite potentials from in situ measured currents. Section 5.5 summarizes our findings and presents some concluding remarks.

5.2 Methodology

In this section, we briefly describe our kinetic simulation approach, and how it is used to construct synthetic data sets used to train and validate inference models, using two regression techniques.

5.2.1 Kinetic simulations

The space plasma parameters considered in our simulations are selected so as to be representative of conditions expected for a satellite in low Earth orbit at altitudes

ranging between 550 and 650 km. This is done by sampling ionospheric plasma parameters using the International Reference Ionosphere (IRI) [23] model in a broad range of latitudes, longitudes, altitudes, and times as shown in Fig. 5.1. The ranges considered for these parameters are summarized in Tab. 5.1. Forty-five sets of plasma parameters approximately evenly distributed in this parameter space are selected as input in simulations, as shown in numbered squares in Fig. 5.1. The three-dimensional PIC code PTetra [19, 20] is used to simulate probe currents in this study. In the simulations, space is discretized using unstructured adaptive tetrahedral meshes [21, 22]. Poisson's equation is solved at each time step using Saad's GMRES sparse matrix solver [91] in order to calculate the electric field in the system. Then, electron and ion trajectories are calculated kinetically using their physical charges and masses self consistently. The mesh for the m-NLP and the simulation domain illustrated in Fig. 5.2, is generated with GMSH [22]. The needle probe used in the simulation has a length of 25 mm and a diameter of 0.51 mm, as those on the NorSat-1. The needle probe is attached to a 15 mm long and 2.2 mm diameter guard which is biased to the same voltage as the probe. The outer boundary of the simulation domain is closer to the probe on the ram side, and farther on the wake side, as shown in Fig. 5.2. The simulations are made using two different domain sizes depending on the Debye length of the plasma. For plasma density below 2×10^{10} m^{-3} corresponding to a Debye length of 1.9-7.2 cm, a lager domain is used. For plasma density above 2×10^{10} m^{-3}, corresponding to a Debye length of 0.2-2.2 cm, a smaller domain with finer resolution is used. The simulation size, the resolution, the number of tetrahedra, and the corresponding Debye length are summarized in Tab. 5.2. There is overlap between the two simulation domains for simulations with Debye lengths around 2 cm. No obvious difference was found in the simulated currents, indicating that simulation results from both domains are consistent in the transition range. Simulation results from both domains are included when training the regression models. All simulations are run initially with 100 million ions and electrons, but these numbers vary through

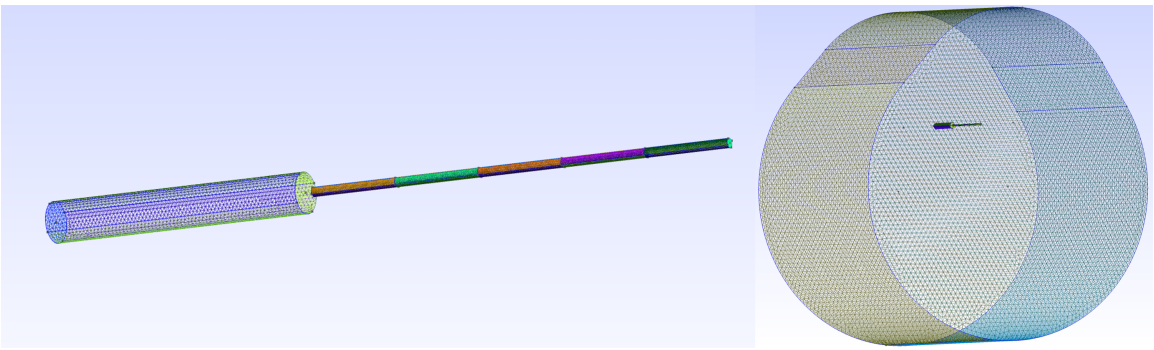

Figure 5.2: Illustration of a m-NLP geometry (left), and the simulation domain (right). The needle probe has a length of 25 mm and radius of 0.255 mm, with a guard of 15 mm in length and 1.1 mm in radius. The ram flow is from the top of the simulation domain and is assumed to be 7500 m/s.

a simulation, due to particles being collected, leaving, or entering the domain. In the simulations, the probe is segmented into five segments of equal length, making it possible to estimate a rough distribution of the current along its length. The current used to build regression models is a sum of the currents of the five different segments. The orbital speed of the satellite is assumed to be fixed at 7500 m/s in the simulations, with a direction perpendicular to the probe. For the voltages considered, probes are expected to collect mainly electron currents. For simplicity, only two types of ions are considered in the simulation, O^+ and H^+ ions, and no magnetic field is accounted for, which is justified by the fact that the Larmor radius of the electron considered is much larger than the radius of the probe.

5.2.2 Synthetic solution library

In order to assess the inference skill of a regression model, a cost function is defined with the following properties: i) it is positive definite, ii) it vanishes if model inferences agree exactly with known data in a data set, and iii) it increases as inferences deviate from actual data. The cost functions used in this work are: the root mean square error,

$$RMS = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} (Y_{mod_i} - Y_{data_i})^2}, \tag{5.2}$$

Table 5.1: Spatial and temporal parameters used to sample ionospheric plasma conditions in IRI, and the corresponding ranges in space plasma parameters.

Environment and plasma conditions	Parameter range
Years	1998 2001 2004 2009
Dates	Jan 4 Apr 4 Jul 4 Oct 4
Hours	0-24 with increment of 8 hours
Latitude	-90° - $+90^\circ$ with increment of 5°
Longitude	0° - -360° with increment of 30°
Altitude	550-650 km with increment of 50 km
Ion temperature	0.07-0.16 eV
Electron temperature	0.09-0.25 eV
Effective ion mass	2-16 amu
Density	$2 \times 10^9 - 1 \times 10^{12} \mathrm{m}^{-3}$

Table 5.2: Parameters used in the two simulation domains are listed. The first two columns give the distances between the probe to the outer boundary on the ram side (D_{ram}), and the wake side (D_{wake}) respectively, followed by the simulation resolutions at the probe, guard, and the outer boundary. The number of tetrahedra used in the simulations is in the order of millions. The corresponding range in Debye lengths is also listed.

D_{ram}	D_{wake}	Probe resolution	Guard resolution	Boundary resolution	Tetrahedra	Debye length
3.5 cm	7 cm	51 μm	220 μm	2 mm	2.5 M	0.2-2.2 cm
30 cm	40 cm	51 μm	220 μm	1 cm	1.7 M	1.9-7.2 cm

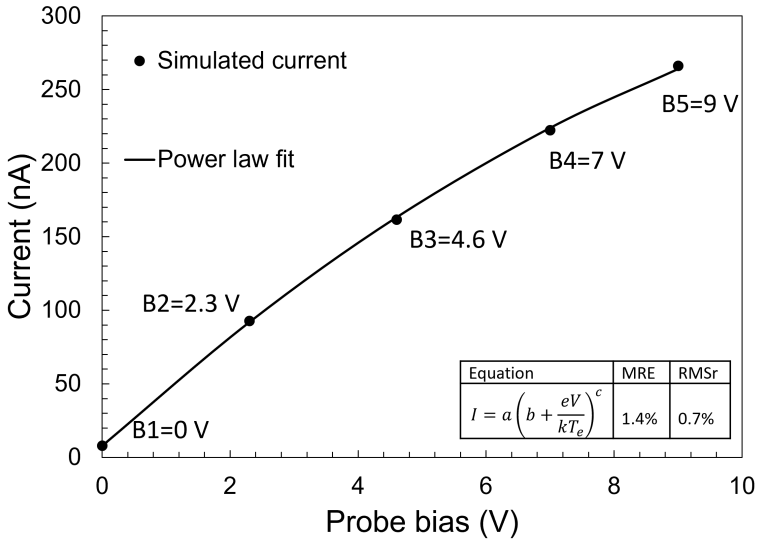

Figure 5.3: Comparison between calculated currents from PIC simulations, and fitted values using Eq. 5.6, assuming a density of 2×10^{10} m^{-3}, an effective mass of 8 amu, an electron and ion temperatures of 0.15 and 0.12 eV respectively, corresponding to point 16 in Fig. 5.1. The fitting errors in the figure are calculated over all 45 sets of plasma conditions using Eq. 5.3 and 5.5.

the root mean square relative error

$$RMSr = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} \frac{\left(Y_{mod_i} - Y_{data_i}\right)^2}{Y_{mod_i}^2}}, \tag{5.3}$$

the maximum absolute error

$$MAE = \max\left\{\left|Y_{mod} - Y_{data}\right|\right\}, \tag{5.4}$$

and the maximum relative error

$$MRE = \max\left\{\left|\frac{Y_{mod} - Y_{data}}{Y_{mod}}\right|\right\}, \tag{5.5}$$

where Y_{data} and Y_{mod} represent respectively known and inferred plasma parameters, and N_{data} is the total number of data points.

For each set of plasma conditions corresponding to a square in Fig. 5.1, 5 simulations are made assuming 5 probe voltages with respect to background plasma, and

the simulated currents vs probe voltage are fitted analytically with:

$$I = a\left(b + \frac{eV}{kT_e}\right)^c, \tag{5.6}$$

where a, b, and c are adjustable fitting parameters. The MRE calculated for all 45 fits is 1.4%, and the RMSr is 0.7%, which shows excellent agreement with simulated collected currents. A comparison between fitted and computed currents is shown in Fig. 5.3. The NorSat-1 m-NLP probes fixed biases V_b are +10, +9, +8, and +6 V, and the probe voltage with respect to background plasma is given by the sum of the spacecraft floating potential plus the probe bias $V = V_f + V_b$. In simulations, probe currents are calculated for voltages with respect to background plasma in the range between 0 to 9 volts are considered as shown in Fig. 5.3. Considering the probe bias voltages V_b given above, probe currents can be determined, corresponding to arbitrary floating potentials between -1 V and -6 V. A synthetic solution library is created for randomly distributed spacecraft floating potentials in the range between -1 and -6 V with corresponding currents obtained by interpolation using Eq. 5.6 with the fitting parameters computed for each of the 45 cases considered. The result is a synthetic solution library consisting of four currents collected by the four needle probes at the four different bias voltages, for 160 randomly distributed spacecraft potentials in the range between -1 V to -6 V for each of the 45 sets of plasma parameters. In each entry of the data set, these four currents are followed by the electron density, the spacecraft potential the electron and ion temperatures, and the ion effective mass as listed in Tab. 5.3. The resulting solution library consisting of $45 \times 160 = 7200$ entries is then used to construct a training set with 3600 randomly selected nodes or entries, and a validation set with the remaining 3600 nodes. The cost functions reported in what follows, used to assess the accuracy of inferences, are all calculated from the validation data set unless stated otherwise.

Table 5.3: Example entries of the synthetic data set, with currents I_1, I_2, I_3, and I_4 calculated using Eq. 5.6, and V_b set to 10, 9, 8, and 6 V, respectively. The floating potential V_f is selected randomly in the range of -1 to -6 V, and the probe voltages with respect to background plasma are given by $V = V_b + V_f$. The coefficients, a, b and c are obtained from a nonlinear fit of the simulated currents using Eq. 5.6. The first and second entries correspond respectively to points 16 and 21 in Fig. 5.1.

$I_1(nA)$	$I_2(nA)$	$I_3(nA)$	$I_4(nA)$	$V_f(V)$	$n_e(\mathrm{m}^{-3})$	T_e(eV)	T_i(eV)	$\mathrm{m_{eff}}$(amu)
233	208	183	129	-2.50	2×10^{10}	0.15	0.12	8
596	533	467	323	-2.93	5×10^{10}	0.07	0.07	4

5.2.3 Multivariate regression

The next step is to construct a multivariate regression model that maps the currents to the corresponding plasma conditions in the solution library. In a complex system where the relation between independent variables and dependent variables cannot readily be cast analytically, multivariate regressions based on machine learning techniques are powerful alternatives to construct approximate inference models. In this approach, the model must be capable of capturing the complex relationship between dependent and independent variables. Once the model is trained using the training set, it can then be used to make inferences for cases not included in the training data set. In this work, two multivariate regression approaches are used to infer plasma parameters: the Radial Basis Function and Feedforward Neural Networks. The models are trained by optimizing their cost function on the training data set, and then applied to the validation data set to calculate the validation cost function without further optimization. The use of a validation set is to avoid "overfitting" because there are certain limitations on the refinement of a model on a training set, such that further improvement of model inference skill in the training set will worsen the model inference skill in the validation set. A good model is one with the right level of training so as to provide the best inference skill in the validation set.

Radial basis function

Radial basis function (RBF) multivariate regression is a simple and robust tool used in many previous studies to infer space plasma parameters using a variety of instruments with promising results [30–32, 112]. A general expression for RBF regression for a set of independent n-tuples $\bar{X}$ and corresponding dependent variable Y is given by:

$$Y = \sum_{i=1}^{N} a_i G\left(\|\bar{X} - \bar{X}_i\|\right). \tag{5.7}$$

In general, the dependent variable Y can also be a tuple, but for simplicity, and without loss of generality, we limit our attention to scalar dependent variables. In Eq. 5.7, the $\bar{X}_i$ represents the N centers, G is the interpolating function, and the a_i are fitting collocation coefficients which can be determined by requiring collocation at the centers; that is, by solving the system of linear equations

$$\sum_{i=1}^{N} a_i G(\|\bar{X}_k - \bar{X}_i\|) = Y_k \tag{5.8}$$

for $k = 1, ..., N$. Here, the dependent variable Y corresponds to the physical parameter to be inferred, and the independent variable $\bar{X}$ is a 4-tuple corresponding to the currents or the normalized currents from the m-NLPs depending on which physical parameters are being inferred. There are different ways to distribute the centers in RBF regression. One straightforward approach is to select centers from the training data set, and evaluate the cost function over the entire training data set for all possible combinations of centers, then select the model which yields the optimal cost function. For this approach, the number of combinations required for $\mathcal{N}$ data points and N centers is given by

$$\binom{\mathcal{N}}{N} = \frac{\mathcal{N}!}{N!(\mathcal{N}-N)!}. \tag{5.9}$$

This, of course, can be prohibitively large and time-consuming for a large training data set or using a large number of centers. An alternative strategy is to successively

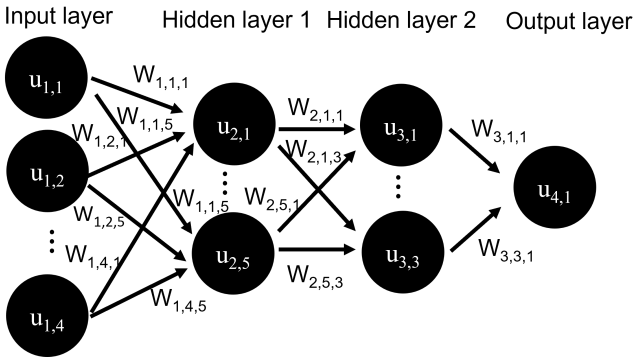

Figure 5.4: Schematic of a feedforward neural network.

train models with randomly selected small subsets of the entire training data set using the straightforward approach, while calculating the cost function on the full training set, and then carrying the optimal centers from one iteration to the next. This "center-evolving strategy" is very efficient in finding near-optimal centers for large training data sets and has proven to be as accurate as the straightforward extensive approach. The RBF models here follow this procedure. Different G functions and cost functions are tested, and only the models that yield optimal results are reported in this paper.

Feedforward neural network

The second multivariate regression approach is a Feedforward neural network as illustrated in Fig. 5.4. This consists of an input layer, hidden layers, and an output layer. Each node j in a given layer i in the network is assigned a value $u_{i,j}$, and the node in the next layer $i+1$ are "fed" from numerical values from the nodes in the previous layer according to

$$u_{i+1,k} = f\left(\sum_{j=1}^{n_i} w_{i,j,k} u_{i,j} + b_{i,k}\right), \tag{5.10}$$

where $w_{i,j,k}$ are weight factors, $b_{i,j}$ are bias terms, and f is a nonlinear activation function [26]. In this work, the input layer neurons contain the four-needle probe currents or normalized currents depending on the physical parameter to be inferred,

whereas the output layer contains one physical parameter. The number of hidden layers and the number of neurons in the hidden layers are adjusted to fit the specific problem, and attain good inference skills. The Feedforward neural network is built using TensorFlow [27] with Adam optimizer [113], and using the ReLU activation function defined as $f(x) = \max(0, x)$. The input variables are normalized using the `preprocessing.normalization` TensorFlow built-in function which normalizes the data to have a zero mean and unit variance. The structure of the network will be described later when presenting model inferences.

5.3 Assessment with synthetic data

In this section, we assess our models using synthetic data, which allows us to check the accuracy, and quantify uncertainties in our inferences. A consistency check strategy is also introduced to further assess the applicability of our models.

5.3.1 Density inference

The density can be inferred using Eq. 5.1 which can be rewritten as

$$\frac{n_e}{T_e^{\beta-\frac{1}{2}}} = \sqrt{\frac{\pi^2 m_e}{2A^2 e^3}} \left(\frac{I_1^{\frac{1}{\beta}} - I_2^{\frac{1}{\beta}}}{V_1 - V_2} \right)^{\beta}. \tag{5.11}$$

In this equation, subscripts 1 and 2 indicate different probes. A special case of this equation was first proposed by Jacobsen, assuming an infinitely long probe, for which $\beta = 0.5$, resulting in

$$n_e = \sqrt{\frac{\pi^2 m_e}{2A^2 e^3}} \sqrt{\frac{I_1^2 - I_2^2}{V_1 - V_2}}, \tag{5.12}$$

which gives an expression for the electron density, independently of the temperature [3]. With currents from more than two probes, the density can be calculated from the slope of the current squared as a function of the bias voltage from a linear least-square fit of all probes [3]. This will be referred as the "Jacobsen linear fit" (JLF) approach.

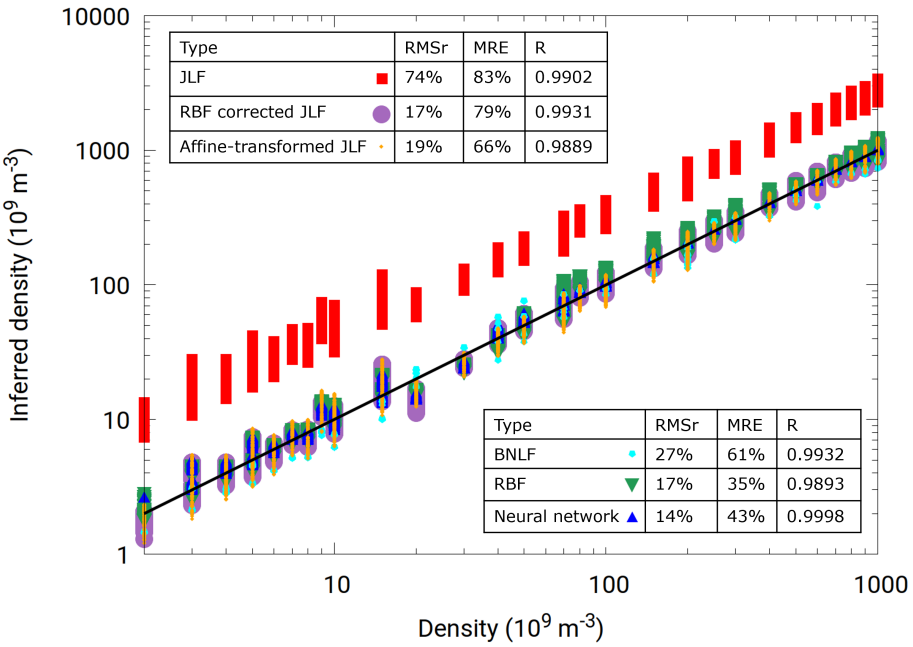

Figure 5.5: Correlation plot for the density inferences made with different techniques applied to our synthetic validation set. The Pearson correlation coefficient R is calculated using the inferred densities and the density used in the simulation. Black line represent idealized perfect correlation line.

On the other hand, the $\beta = 0.5$ assumption requires that the needle probe be very long compared to the Debye length, which is in general not satisfied for NorSat-1 satellite. As a consequence, when this method is applied to the solution library, the inferred density is typically three times larger than the actual density as shown with red boxes in Fig 5.5. Despite this offset, the high Pearson correlation coefficient R shows that inferences made with this method can be significantly improved with a simple affine transformation. The best results are obtained by applying an affine transformation to the log of the JLF inferred density as in:

$$\ln(n_e') = a \ln(n_e) + b. \tag{5.13}$$

In this equation, the density n_e is first obtained using the JLF method, then an affine transformation is used to calculate the inferred density n_e'. The affine transformation coefficients a and b are obtained from a least-squares fit of the log of these densities, to those in the training data set. The fitting coefficients in this case, $a = 1.13261$ and

$b = -4.82735$, are then used to perform an affine transformation on the validation data set, leading to a significant improvement in RMSr from 74% to 19%, and in MRE from 83% to 66% compared to densities inferred from the JLF approach, as shown in Fig 5.5. RBF regression can also be used to correct JLF density. This is done by using RBF to approximate the discrepancy between the densities used in the simulations and the ones inferred with JLF. This correction is then used to improve the accuracy of the inferred density obtained with JLF method. Using the four currents as input variable $\bar{X}$, by minimizing the MRE, using $G(x) = |x|$, and 5 centers, the RBF corrected JLF density yields an RMSr of 17 % and a MRE of 79%. The cost functions of the two methods are comparable, but an obvious advantage of using an affine transformation is its simplicity.

Barjatya's nonlinear least square fit method is also assessed using our synthetic data set. The original method was applied to sweep mode measurements, to obtain the electron temperature and the satellite potential from currents in the ion saturation region and electron retardation region currents, before fitting the density and β from the electron saturation region [10]. This is however not possible with fixed bias probe measurements considered here. On NorSat-1, four currents are measured simultaneously, by four probes at different fixed bias voltages, all in the electron saturation region. A similar approach can nonetheless be applied in our case, using a nonlinear fit to the currents, with the density, the electron temperature, the satellite potential, and β, as fitting parameters. As shown by Barjatya and Merritt [13], however, it is difficult to infer the temperature using this approach, owing to the weak dependence of collected currents on the electron temperature (see Eq. 5.11). A solution, proposed in [4, 13], then consists of estimating the electron temperature from other measurements, or from the IRI model, and perform the fit for the remaining three parameters. This simplification is justified by the fact that, following this procedure, a 50% error in the temperature, still produces acceptable results for the other param-

eters [13]. Thus in this study, we assume a fixed electron temperature ($\sim$ 2000 K), which is in the middle of the temperature range considered in the simulations, and fit 4-tuples of currents using potentials, densities, and β values as fitting parameters. This will be referred as the "Barjatya nonlinear fit" (BNLF) approach. The Python 3 `differential.evolution` package is used to do the nonlinear fit with an evolution strategy of 'best2exp', with a tolerance of tol=0.001. In the fits, the upper and lower bounds for the density, the potential and the β value are 1×10^9 to 1×10^{12} m^{-3}, -6 to -1 V, and 0.5 to 1, respectively. The potential lower bound of -6 V is needed to ensure that the values under exponent in Eq. 5.1 are positive. We obtain 3600 fits for each of the 3600 entries of four currents in our validation data set. The fit minimizes RMSr as the cost function, and the overall RMSr calculated using Eq. 3.5 for the 3600×4 currents is 0.02 %, and only 0.26% of the points have relative errors larger than 1%. The resulting density inferences have an RMSr of 27 % and a MRE of 61 %, which is better than the densities inferred from the JLF approach, but less accurate than the affine-transformed JLF density. The β values calculated are in the range of 0.75 to 1. The inferred potential using this method is discussed in the next section together with other methods. With only four fitting points, the fit can fail into local minimums instead of the global minimum, thus, the tolerance of the fit must be small. As a result, the nonlinear fits tend to be somewhat time-consuming. with each fit requiring approximately 1 second using an AMD 5800x processor. In comparison, linear fits of the currents square, followed by an affine transformation of the log of the inferred density can be done using fixed formulas, and thus are considerably faster than a nonlinear fit. Regression methods such as RBF or neural network are also numerically very efficient, considering they involve simple arithmetic expressions with pre-calculated coefficients.

Direct RBF regression can be applied to infer density using the four currents as input variables. When constructing an RBF model with $G(x) = |x|$, minimizing MRE, and using 6 centers, the RMSr and MRE calculated on the validation data set

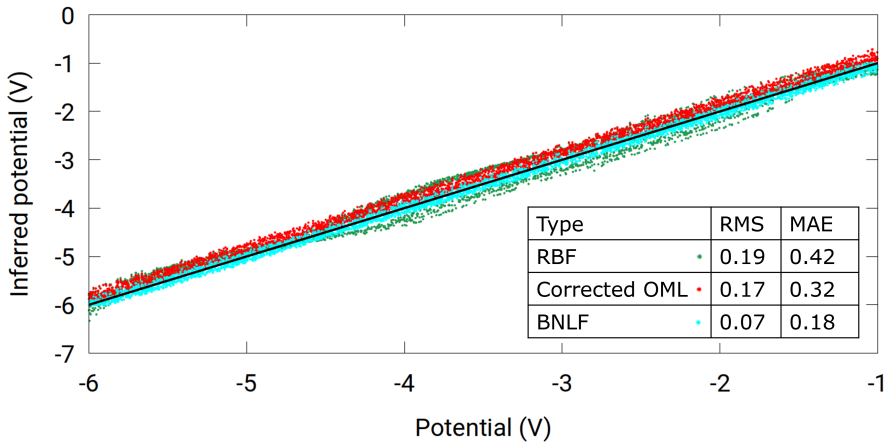

Figure 5.6: Correlation plot obtained for satellite potential inferred with RBF and OML techniques.

are 17% and 35%, respectively. Using a neural network with 4 nodes in the input layer, 14 nodes and 12 nodes in two hidden layers, and 1 node in the output layer, results in a 14% RMSr and 43% MRE for the inferred densities. This is calculated using TensorFlow with ADAM optimizer with a learning rate of 0.005 and an RMSr as a cost function. The input layer is normalized to have a zero mean and unit variance, while the output layer is normalized by dividing the largest density. The densities calculated using the synthetic solution library, as well as the cost function are shown in Fig. 5.5. Compared to the other density models considered, straightforward RBF yields the smallest MRE, thus it is the preferred model to infer density in this work. However, the affine-transformed JLF method enables density inferences with accuracy comparable to those of more complex approaches. This simple and practical technique should be of interest in routine data analysis.

5.3.2 Potential inference

The floating potential of the spacecraft can also be inferred using the OML equation, by rewriting equation 5.1 as:

$$V_f \approx V_f + \frac{kT_e}{e} = \frac{V_2 I_1^{\frac{1}{\beta}} - V_1 I_2^{\frac{1}{\beta}}}{I_2^{\frac{1}{\beta}} - I_1^{\frac{1}{\beta}}} = \frac{V_3 I_2^{\frac{1}{\beta}} - V_2 I_3^{\frac{1}{\beta}}}{I_3^{\frac{1}{\beta}} - I_2^{\frac{1}{\beta}}}. \tag{5.14}$$

In this equation, the subscripts 1,2, and 3 refer to different probes, thus there must be at least three probes in order to solve for β. The bias voltages of the probes and their corresponding collected currents are known from measurements, thus β can be solved using a standard root finder. Given β, equation 5.14 then provides a value for $V_f + \frac{kT_e}{e}$. In this expression, $\frac{kT_e}{e}$ is the electron temperature in electron-volt, which in the lower ionosphere at mid latitudes, is of order 0.3 eV or less. Thus, considering that $\frac{kT_e}{e}$ is generally much smaller than satellite potentials relative to the background plasma, any of the two terms in the right side of Eq. 5.14 provides a first approximation of V_f [31]. This will be referred to as the "corrected OML" approach. This equation works well when it is applied to the synthetic solution library with a MAE of 0.3 V calculated using currents collected with probe biases of 10, 9, and 8 volts probes. The error of 0.3 V is likely due to the maximum electron temperature of 0.3 eV considered in the simulations. The β values calculated in the synthetic solution library is in the range of 0.75 to 1. It is also possible to build a model to infer floating potentials directly using RBF regression. In this case, currents are normalized by dividing every current by their sum, in order to remove the strong density dependence on the currents. Using $G(x) = |x|$, and 5 centers, and minimizing MAE, the calculated MAE on the validation data set is 0.4 V. The inferred satellite potential from the BNLF approach has an RMS of 0.07 V, and a MAE of 0.18 V, which proves this method to be the most accurate compared to the other methods considered. A correlation plot for potentials inferred using the RBF, corrected OML, and BNLF approaches is shown in Fig. 5.6. All methods show good agreement with values from the synthetic solution library.

5.3.3 Consistency check

In order to further assess the applicability of our inference approaches, we perform the following consistency check. First, RBF models $M1(n_e)$ and $M1(V_f)$) are constructed to infer the density and satellite potential using 4-tuple currents from our synthetic

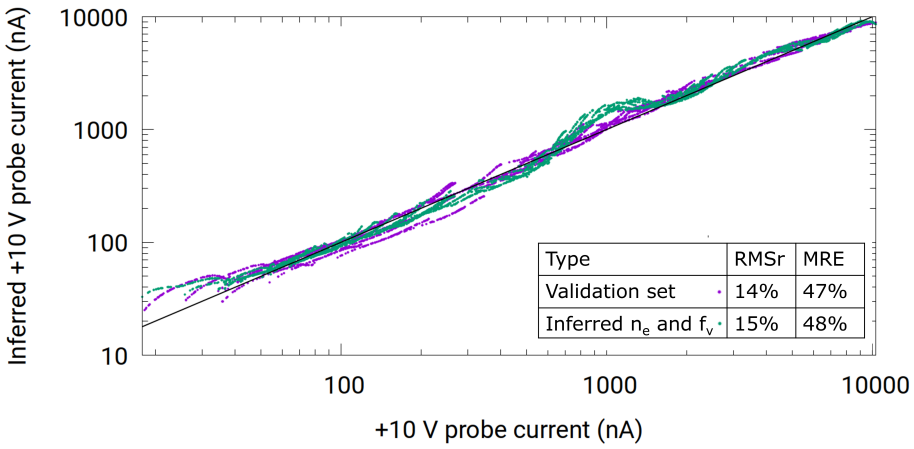

Figure 5.7: Correlation plot of inferred +10 V probe current against +10 V probe current from the synthetic data set is presented. The calculated +10 probe currents in purple curve is calculated using the validation data set, while the green curve is calculated using inferred densities and floating potentials from RBF regression.

data set. A second model ($M2$) is constructed to infer collected currents from densities and floating potentials in our synthetic data set. Since we are not able to infer temperatures from the currents, the temperature is not included in $M2$. Consistency is then assessed in two steps, by i) using currents from synthetic data and models $M1(n_e)$ and $M1(V_f)$ to infer densities and floating potentials, and ii) applying models $M2$ to these inferred values, to infer back collected currents. RBF density and floating potential inferences are used in $M1(n_e)$, and $M1(V_f)$ as described in sec. 5.3.1 and 5.3.2. RBF is also used in $M2$ with $G(x) = \sqrt{1 + x^{2.5}}$, and minimizing RMSr with 5 centers. With perfect inference models, the results for these back-inferred currents, should agree exactly with the starting currents from synthetic data. Variances between back-inferred and simulated currents in the synthetic data are presented as indicative of the level of confidence in our regression techniques. The correlation plot in Fig. 5.7, shows back-inferred currents (green) calculated for a probe with 10 V bias against known currents from synthetic data. For comparison, the figure also shows the correlation between directly inferred currents (purple) when model M_2 is applied to densities and floating potentials in the synthetic data set. Both back-inferred and directly inferred currents are in excellent agreement with known currents from syn-

thetic data, with comparable metric skills of $\simeq 15\%$ and $\simeq 48\%$ for the RMSr and the MRE, respectively. Considering that errors are compounded between the first and second models for the back-inferred currents, the nearly identical metric skills in Fig. 5.7 is seen as confirmation of the validity of our regression models.

5.4 Application to NorSat-1 data

In this section, we apply our density and potential inference models constructed with synthetic data, to in situ measurements made with the m-NLP on the NorSat-1 satellite. The NorSat-1 currents were obtained from a University of Oslo data portal [6]. The epoch considered corresponds to one and a half orbit of the satellite starting at approximately 10:00 UTC on January 4, 2020. We start with a comparison of simulated and measured currents to verify that our simulated currents are in the same range as those of measured in situ currents. Inferences made with RBF, neural network, BNLF , corrected OML, and the two corrected JLF approaches constructed in 5.3.1, are also presented.

5.4.1 Measured in-situ, and simulated currents

The relevance of the space plasma parameter range considered in the simulations, to NorSat-1, is assessed in Fig. 5.8, by plotting currents collected by the +9 V probe against that collected by the +10 V, from both synthetic data, and in situ measurements. The close overlap, and the fact that the range of in situ measurements is within the range of simulated currents, indicates that the physical parameters selected in the simulations, are indeed representative to conditions encountered along the NorSat-1 orbit.

The current measurement resolution for the NorSat-1 m-NLP probes is approximately 1 nA [6]. The noise level from the environment, however, is estimated to be of order 10 nA. In what follows, darker colors are used to represent inferences made

using currents above 10 nA, and lighter colors are used to represent inferences using currents between 1 to 10 nA. This is done by filtering out all data that contain a current that is below 10 nA or 1 nA in any of the four probes. A word of caution is in order, however, for inferences made from these lower currents, as a conservative estimate of the threshold for sufficient signal-to-noise ratios, is approximately 10 nA. This lower bound current is supported by a consistency check made with models 1 and 2 described in Sec. 5.3.3, and presented below in Sec. 5.4.3.

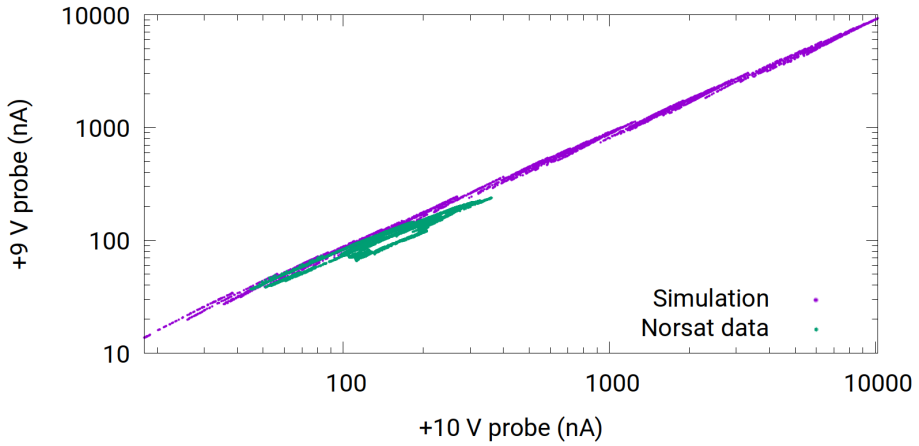

Figure 5.8: Correlation plot between currents collected by the +9 V and the +10 V probes for both NorSat-1, and synthetic data.

5.4.2 Density and satellite potential inference

Our models, trained with synthetic data as described in Sec. 5.3, are now applied to infer plasma densities and satellite potentials from in situ measured currents, for the time period considered. The results obtained with the different models presented in Sec. 5.3 are shown in Fig. 5.9 for the inferred densities, satellite potentials, and measured currents collected by the four probes. The position of the satellite relative to the Earth and the Sun given by the solar zenith angle, is also plotted in the figure. For example, a small solar zenith angle means that the satellite is near the equator on the dayside.

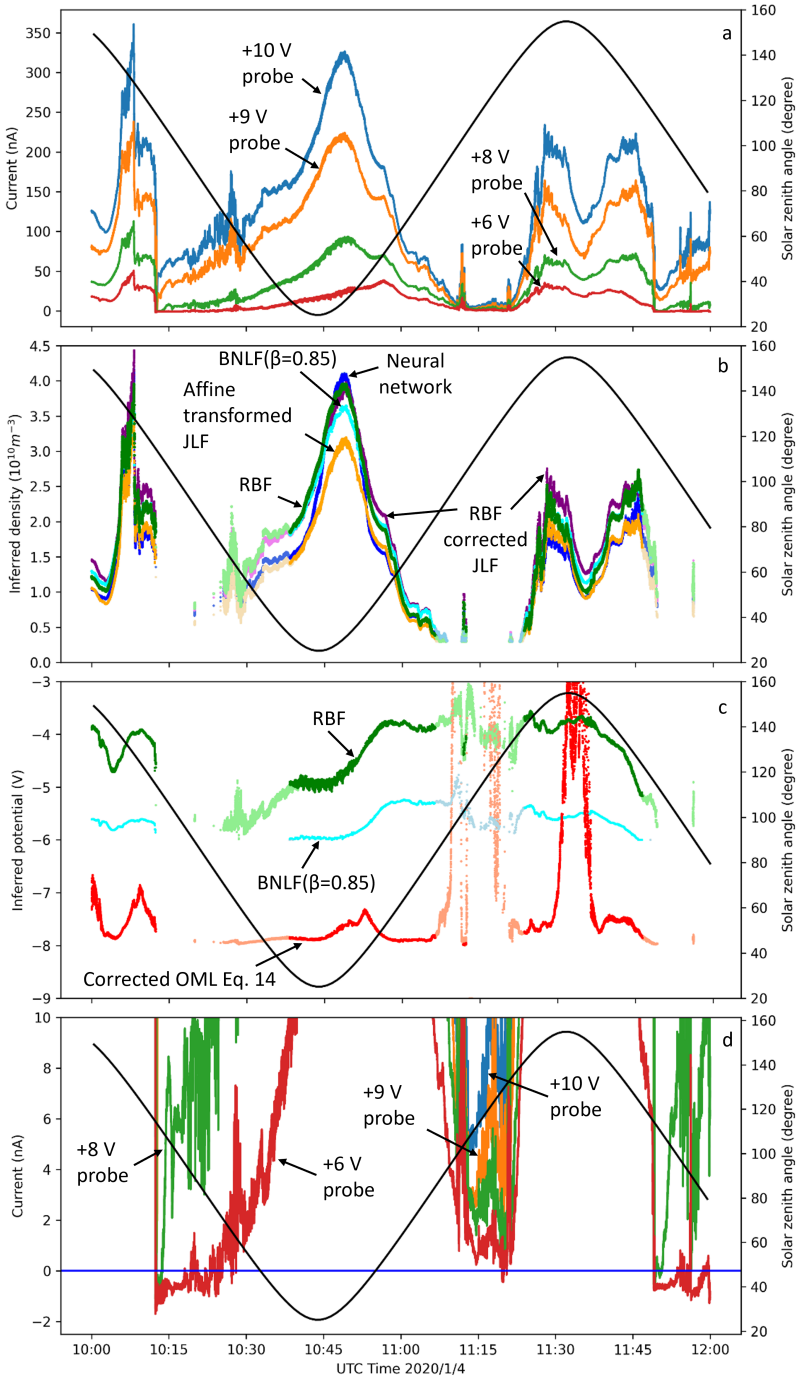

Figure 5.9: Illustrations of NorSat-1 collected currents considered in this study in panel a, inferred densities in panel b, inferred potentials in panel c, and the NorSat-1 current near 0 A in panel d. The solar zenith angle is also plotted against the secondary axis. Curves in darker colors are from model inferences using data above 10 nA, whereas those in lighter colors show inferences using data with currents between 1 nA and 10 nA.

Fitting the in-situ data with the BNLF method using the four measured currents is challenging. The +8 V probe currents are often slightly lower than expected for a downward concavity of I as a function of V_b, and tend to produce an upward concavity with β larger than 1. In this case, fitting the 4-tuples of currents using Eq. 5.1, for the density, the floating potential and β, with a specified electron temperature is not practical. Thus we used a fixed β value of 0.85, and fit only density and potential to the 4-tuples of currents using Eq. 5.1. This choice for the value of β is justified by the fact that it produces the best inferences when applied to synthetic data, with an RMS error of 0.39 V for the floating potential, and an RMSr error of 27 % for the density. Based on comparisons made with our synthetic data sets, the use of a fixed β value results in a small loss in the inference accuracy for the satellite potential, but the accuracy of the inferred density is the same as when β is included as a third fitting parameter. Using the fixed values of 0.172 eV for the electron temperature, and 0.85 for β, the RMSr error in the fits of the measured in situ currents, is 9%. The resulting inferred densities and satellite potentials are shown in Fig. 5.9. For reasons mentioned above, it is clear that no satellite potential below the fitting lower bound of -6 V can appear in the plot. On a practical note, an advantage of computing BNLF inferences with fixed temperature and β value, is that the nonlinear fit is made for only two fitting parameters (n_e and V_f), which results in faster convergence rates, compared to fits made with 4 or 3 adjustable parameters. In our calculations, for example, the convergence rate is 8 times faster with two, compared to three fitting parameters.

The densities shown in Fig. 5.9 panel b are obtained using the five density inference methods mentioned in Sec. 5.3.1. At 10:45, the neural network density, the RBF corrected JLF density, the RBF density, and the BNLF density ($\beta = 0.85$) overlap nicely, while the affine transformed JLF density is smaller than other inferred densities, particularly near the density maxima. The density inferences nonetheless qualitatively agree with each other. Using the +10, +9, and +8 NorSat-1 probe cur-

rents and Eq. 5.14, the inferred satellite floating potential is about -8 V for most of the data range considered in this study as shown in Fig. 5.9 panel c. This is in stark contradiction with observations in Fig. 5.9 panel d, which shows that the +6 V biased probe collects net positive electrons during most of the period considered. Also, there are periods between 10:15 to 10:30, and after 11:45 when the + 6V probe collects ion current(negative), indicating drops in the satellite potential below -6V. The poor performance of Eq. 5.14 to infer the satellite potential here, results from the fact that Eq. 5.14 yields erratic values of β ranging from 0.3 to 1.2. Attempts have also been made to approximate the satellite potential with Eq. 5.14 using a fixed value of 0.58 and 0.78 for β, also resulting in satellite potentials in the -8 V range, and no improvement was found. This failure to produce acceptable values of the satellite potential clearly shows that this generalized OML approximation in Eq.5.14 does not provide a sufficiently accurate approximation for the currents collected by the NorSat-1 probes.

The RBF inferred floating potential shown in Fig. 9, is within -4 and -6 V, which is consistent with the observation that the +6 V probe collects electrons during most of the time period considered. Interestingly, the inferred satellite potential using currents between 1 and 10 nA (light color) is seen to join smoothly with the darker color inferences, and to decrease below -6 V around 10:25, which is consistent with the observation that during that time the +6 V probe no longer collects electron current. The floating potentials inferred from the BNLF model are systematically lower than those from RBF, and they also fit within the acceptable range for the satellite potentials. The two potentials have otherwise a very similar time dependences. The +6 V probe collects zero net current near 10:25 in panel d. The BNLF potential is bounded by the fitting lower limit of -6 V at these ranges, as opposed to RBF with which inferences are made without imposing an upper or lower bound. The currents collected by the probes are determined mostly by the density and the satellite po-

tential, and to a lesser extent, by the electron temperature. In Fig. 5.9, the density and floating potential are seen to peak at around 10:45 and 11:00 respectively. The currents from the +8, +9, and +10 V probes (green, orange, and blue) in panel a peak at around 10:45, coinciding with the peak in the plasma density at this time. Then, as time goes forward to 11:00, the currents of the three probes decrease, also coinciding with a decrease in plasma density. However, the +6 V probe (red) current is increasing during these times, possibly due to an increase in floating potential. This increase is captured in the RBF and BNLF inferred potential, but not in the one derived from corrected OML. Another observation is that the inferred floating potential decreases significantly at 10:15, as the satellite crosses the terminator. On NorSat-1, the negative terminals of the solar cells are grounded to the spacecraft bus while the positive side is facing the ambient plasma [114]. A likely explanation for the potential drop is that the solar cells facing the ambient plasma get charged positively and suddenly start collecting more electrons upon exiting solar eclipse. This would agree with findings reported by Ivarsen et al. [114].

5.4.3 Consistency check

In the absence of accurate and validated inferred densities and satellite potentials from NorSat-1 data, it is not possible to confidently ascertain to what extent the inferences presented above are accurate. As an alternative, we proceed with a consistency check, following the same procedure as presented in Sec. 5.3.3 with synthetic data, but using measured currents as input. This is done by first applying models $M1(n_e)$ and $M1(V_f)$) trained with synthetic data, to infer floating potentials and densities from measured currents. Then $M2$ (also trained with synthetic data) is used to infer currents from the $M1$ - inferred floating potentials and densities. If the models constructed from the synthetic data also apply to NorSat-1 data, the inferred currents should closely reproduce the measured NorSat-1 currents. A correlation plot of inferred against measured currents is shown in Fig. 5.10 for the +10 V probe. In

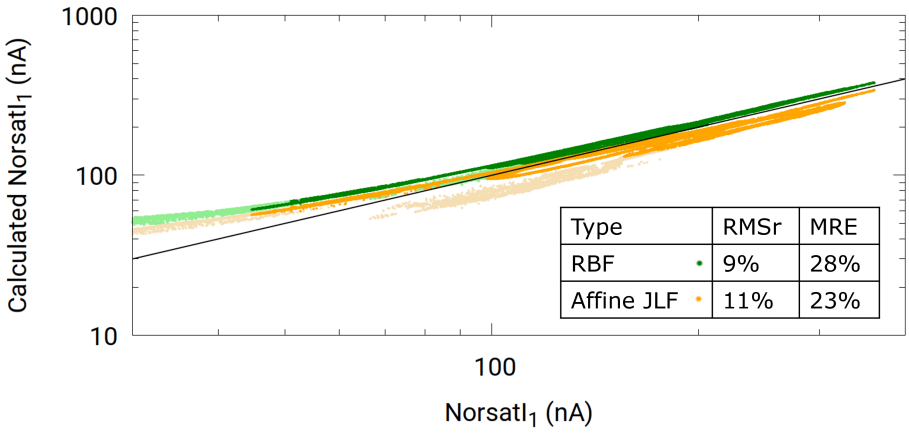

Figure 5.10: Consistency check is performed in the in situ data following the same procedure as in the synthetic data set. Both models 1 and 2 are trained with our synthetic data, and applied to currents from the +10 V probe on NorSat-1. Darker colors refer to inferences made with currents above 10 nA, while lighter colors refer to inferences obtained with currents between 1 and 10 nA.

this plot, the orange and green curves show back-inferred currents obtained with the RBF $M2$ model. For the orange curve (Affine JLF), the density used as input in $M2$ is obtained with the affine transformed JLF method. For the green curve (RBF), the density used as input in $M2$ is obtained with RBF density, while in both cases, the floating potentials are obtained with the $M1(V_f)$ model from RBF regression. The parts in lighter color are obtained using data with a 1nA filter, whereas the darker color parts are obtained using data with currents above 10 nA. While the graph only shows currents above 30 nA, the 1 nA filter curve extends to the left down to about 5 nA, however, these calculated +10 volt probe currents plateau in this range and are far from the measured currents. This behavior is likely due to noise levels of about 10 nA, thus extra caution should be taken when using model inferences for data below 10 nA. The RMSr calculated for the 10 nA NorSat-1 current using direct RBF density as $M1(n_e)$ is 9%, and the MRE is 28 %, whereas these numbers for the affine transformed JLF densities are 11 % and 23 %, respectively. The calculated +10 V probe currents based on RBF regression and affine transformed JLF method nicely follow the measured +10 volt probe current except for a small increase in the

variance at lower currents, thus indicating that our model constructed with synthetic data set should be applicable to in situ data.

5.5 Conclusions

Two new approaches are presented and assessed, to infer plasma and satellite parameters from currents measured with multiple fixed bias needle Langmuir probes. In the first approach, inferences are made with two multivariate regression techniques, consisting of radial basis functions, and neural networks. The second approach relies on a simple affine transformation combined with a technique first proposed by Jacobsen to infer the plasma density. Yet another approach, proposed by Barjatya, et al. is considered, which consists of performing nonlinear fits of measured currents, to an analytic expression involving the density, the floating potential and the exponent β as fitting parameters, while the electron temperature is estimated by other means. In all cases, the accuracy of inferences is assessed on the basis of synthetic data obtained from kinetic simulations made for space-plasma conditions representative of those encountered along the NorSat-1 satellite. In addition to assessments based on synthetic data, a consistency check is presented, whereby densities and satellite potentials inferred from collected currents, are used as input in an inverse regression model to infer currents for one of the probes. The advantage of this consistency check is that it is applicable to both synthetic, and in situ measured currents, and in the latter case, it does not rely on a priori given inferred densities and satellite potentials. Inference consistency checks are made with both synthetic and in situ measured currents, showing excellent agreement.

The density inference methods considered in this study yield excellent results when applied to the synthetic data set. The models constructed with synthetic data are then applied to currents measured by the four m-NLP on NorSat-1. When applied to NorSat-1 data, the Barjatya nonlinear fit approach is modified by assuming a fixed

value for the temperature and β, and carrying the fit with only the electron density and satellite potential as fitting parameters. The density inferences from all methods show good agreement, which suggests that either method should be a significant improvement over the commonly used OML approach based on $\beta = 0.5$. From on our findings, direct RBF and the combination of Jacobsen's linear fit with $\beta = 0.5$ with an affine transformation, appear as being the most promising, and deserving of further study. These two methods provide inferences that are consistent and quantitatively similar, while being relatively simple and numerically efficient. The former yields the lowest maximum relative error when assessed with synthetic data, whereas the latter is the simplest method and produces inferences with comparable accuracy. The spacecraft floating potential is also inferred using RBF regression, a modified OML approach and Barjatya nonlinear fit method. The modified OML inferences are inconsistent with the measurements from NorSat-1 data since it indicates that the satellite potential is below -6V, while measurements indicate that the +6 V probe is collecting electron current. Conversely, spacecraft potentials inferred with RBF regression, and the nonlinear fit approach yield potentials that are consistent with measured currents from the +6 V biased probe, showing that the satellite potential must have been at or above -6 V for most of the one and a half orbital period considered. This failure to produce acceptable values of the satellite potential using Eq. 5.14, and the fact that the Barjatya nonlinear fit approach with n_e, V_f, and β as fitting parameters, results in β values appreciably larger than one, shows that in situ measurements on NorSat-1 generally do not closely follow the empirical expression in Eq. 1.

The analysis presented here has been focused on fixed bias multi-needle Langmuir probes, with the same dimensions as the ones mounted on NorSat-1, to which it has been applied as a case study. We stress, however, that the simulation-regression approach to infer space plasma parameters, is not limited to fixed bias probes or to this

particular configuration of probes. With kinetic solutions capable of reproducing analytic results in conditions when they are valid, and also capable of accounting for more physics, and more realistic geometries than theories, solution libraries, training and validation sets can just as well be constructed for different probes, mounted on satellites, operated in fixed or sweep bias voltage mode. By following standard machine learning procedures, whereby models are trained on a subset of a solution library of known independent and dependent variables, and tested by applying them to distinct subsets, we can estimate uncertainty margins specifically associated with different inference techniques. Another important strength of the proposed simulation-regression approach is that it enables relatively straightforward incremental improvements to a model, by accounting for more physical processes or more detailed geometries; something that would be very difficult to do in a theory. Implementation of regression models and affine transformation of the Jacobsen linear fit model involve simple arithmetic expressions with pre-calculated coefficients and can easily be programmed for onboard processing of low level data. These approaches, however, would require the creation of custom data sets, when applied to a given mission, so as to account for the geometry relevant to the measuring instruments, and the space environment conditions expected along a satellite orbit. This is where the BNLF technique could prove convenient, as it does not rely on the construction of extensive synthetic data sets and training strategies. Custom regression models, however, would require more computational resources, which would necessitate optimization in order to be implemented onboard a satellite. The work presented here is by no means final. The development of improved inference approaches based on simulations and regression techniques will require significantly more efforts, involving collaborations between experimentalists and modelers; an effort well worth doing, considering the cost and years of preparation involved in scientific space missions.

Chapter 6

Conclusion

6.1 Summary

The main goal of this thesis is to take advantage of the computational power available today to do new science. Particle in cell (PIC) simulations are used to simulate the current collected on selected spacecraft instruments in conditions relevant to near-Earth orbit in low and mid latitudes. Machine learning techniques are used to create regression models to interpret the results from simulations. The uncertainties in the regression models are assessed, and the inferences are compared with existing state-of-the-art theoretical models. In addition, a new efficient algorithm for Radial Basis Function multivariate regression is also presented.

In Chapter 1 and Chapter 2, the background and methodology are presented. In Chapter 3, a new algorithm for radial basis function is presented. Speed can be a serious issue when training a model with Radial Basis Functions (RBF) on a large data set. This is particularly true when training is made with a straightforward construction of models for all combinations of N centers among $\mathcal{N} \gg N$ nodes, and selecting the one leading to the highest inference skill, as per a given cost function. This is because the computational cost in such "exhaustive training" increases exponentially with the number of nodes in the training set. In contrast, the fast Center-Evolving (CE) RBF presented here is significantly more efficient, and proves in practice, to be

nearly as accurate as the exhaustive approach. The CE strategy consists of training models on randomly selected smaller subsets of size $\tilde{\mathcal{N}}$ of the training set, with $N < \tilde{\mathcal{N}} < \mathcal{N}$, while retaining the optimal centers from one subset to the next. This in turn, ensures that the cost function never increases from one $\tilde{\mathcal{N}}$-subset to the next. Example applications are presented where differences in CE training times are reduced by approximately two orders of magnitude compared with those of the exhaustive training.

In Chapter 4, a relatively simple design of a segmented flow meter is presented for measuring in situ plasma flow velocities and other space plasma parameters. The response of the flowmeter to space environment is simulated for plasma conditions representative of the ionosphere at mid and low latitudes using PTetra. A synthetic data set consisting of ion currents collected by several segments of the flow meter, and the physical parameters for which they were calculated, is then used to construct a solution library from which inference models can be constructed, using RBF and neural network regressions. Simulation results show that with such a flow meter, it should be possible to infer plasma flow velocities in the direction perpendicular to the ram direction, with uncertainties of 45 m/s or less. Models can also be constructed to infer plasma densities, with a maximum relative error of 23 %. This work is presented as a first assessment and proof of concept for an original design of a simple and robust flow meter.

Finally, in Chapter 5, multi-needle Langmuir instrument onboard NorSat-1 satellite is studied using PIC simulations. Current inference techniques for processing multi-needle Langmuir Probe (m-NLP) data are often based on adaptations of the Orbital Motion-Limited (OML) theory which relies on several simplifying assumptions. Some of these assumptions, however, are typically not well satisfied in actual experimental conditions, thus leading to uncontrolled uncertainties in inferred plasma

parameters. In order to remedy this difficulty, three-dimensional kinetic particle in cell simulations are used to construct a synthetic data set, which is used to compare and assess different m-NLP inference techniques. Using the synthetic data set, regression-based models capable of inferring electron density and satellite potential from 4-tuples of currents collected with fixed-bias needle probes similar to those on the NorSat-1 satellite, are trained and validated. The regression techniques presented enable excellent inferences of the plasma density, and floating potentials. The new inference approaches presented are applied to NorSat-1 data, and compared with existing state of the art inference techniques.

6.2 Contributions

The main accomplishments of my thesis are summarized below:

- The formulation of an original algorithm for Radial Basis Function that is particularly efficient in training using large data sets.
- The design and assessment of a new type of plasma flowmeter using simulation results and regression techniques.
- The creation of simulation-based regression models and their assessment by comparison with state-of-the-art inference techniques for m-NLP instrument.
- The assessment of various state-of-the-art m-NLP inference techniques and simulation-based regression models using NorSat-1 in-situ data.

6.3 Future work

Multi-needle Langmuir probe instrument

- There are studies for Langmuir probes in a magnetic field, but few can be applied under realistic space plasma conditions. This would be particularly

difficult to study with theoretical approaches but it could be a potential future project using a simulation approach. Probe characteristics could be calculated as a function of magnetic field strength and angle, the satellite attitude, as well as the orientation and magnitude of the plasma drift velocity.

- With a satellite moving at approximately 7500 m/s in a magnetic field, there can be potential differences between satellite components resulting from the motional electric fields ($\vec{v} = -\vec{E} \times \vec{B}$). Some probes might have effective potentials with respective to the background plasma, that would be different from the ones in the absent of the magnetic field. This effect might be studied in another project.

- It would be interesting to investigate the NorSat-1 m-NLP instrument probe biases. If the fixed bias probe voltage under a long operation is shifted to a different voltage, it is hard to detect this shift without a calibration. For example, the fact that the + 8 V probe current is often slightly lower than expected for a downward concavity of collected current as a function of bias voltage might be indicating that the real bias voltage of this probe is less than + 8 V.

- There are other cube satellites with identical m-NLP probes as these onboard NorSat-1. The regression models created in this study are ready to be applied to these satellite instruments.

Plasma flowmeter

- The vector method to interpret measurements with symmetry is a very convenient approach that might be applicable to other instruments with symmetry, for example, segmented spherical Langmuir probe.

- Building the flowmeter, testing it in a lab, and eventually using it on a spacecraft would make an interesting extension of this project.

CERBF program

- A similar algorithm to CERBF can be applied to neural networks. The cost function of a neural network model as a function of the various coefficients has different local minimums. When training a model, the neural network coefficients are initially determined using the Monte Carlo method and then refined using, for example, the gradient descent method. However, in such an approach, the coefficients might be trapped in local minimums and never be able to reach the global minimum if the coefficients determined from the Monte Carlo method are far. This is especially true if the training time is short. Therefore, multiple training trials using neural networks with identical inputs might result in different cost functions. A more efficient approach is to restart the training multiple times and select the lowest cost function among the trials. This approach takes advantage of gradient descent and finds the lowest cost function once the coefficients are close to the optimal values.

- In CERBF, randomly selected subsets of the full training set are used to do extensive searches. An alternative would be to select subsets using approaches such as K-clustering or Gaussian-clustering.

Thesis-Related Publication List (as of July 2022)

Journals

1. G. Liu, Marchand, R. (2021): Kinetic Simulation of Segmented Plasma Flow Meter Response in the Ionospheric Plasma, Journal of Geophysical Research: Space Physics, DOI: 10.1029/2021JA029120

2. G. Liu, Marchand, R. (2022): Inference of m-NLP data using Radial Basis Function regression with center-evolving algorithm, Computer Physics Communication, DOI: 10.1016/j.cpc.2022.108497

3. G. Liu, Marholm, S. Eklund, A. Clausen, L. Marchand, R. (2022): m-NLP inference models using simulation and regression techniques, Journal of Geophysical Research: Space Physics, under review, preprint: https://doi.org/10.1002/essoar.10510978.2

Computer program

1. CERBF from https://codeocean.com/capsule/7662817/tree

Bibliography

[1] W. Baumjohann and R. A. Treumann, *Basic Space Plasma Physics (revised edition)*. Imperial College Press, World Scientific Publishing Company, 2012.

[2] H. Hoang, K. Røed, T. A. Bekkeng, J. I. Moen, L. B. Clausen, E. Trondsen, B. Lybekk, H. Strøm, D. M. Bang-Hauge, A. Pedersen, C. D. Nokes, C. Cupido, I. R. Mann, M. Ariel, D. Portnoy, and E. Sagi, "The multi-needle langmuir probe instrument for QB50 mission: Case studies of Ex-Alta 1 and hoopoe satellites," *Space Science Reviews*, vol. 215, 2 2019.

[3] K. S. Jacobsen, A. Pedersen, J. I. Moen, and T. A. Bekkeng, "A new langmuir probe concept for rapid sampling of space plasma electron density," *Measurement Science and Technology*, vol. 21, 8 2010.

[4] H. Hoang, K. Røed, T. A. Bekkeng, J. I. Moen, A. Spicher, L. B. N. Clausen, W. J. Miloch, E. Trondsen, and A. Pedersen, "A study of data analysis techniques for the multi-needle langmuir probe," *Measurement Science and Technology*, vol. 29, p. 065 906, 6 Jun. 2018.

[5] S. Basu, S. Basu, E. MacKenzie, W. R. Coley, J. R. Sharber, and W. R. Hoegy, "Plasma structuring by the gradient drift instability at high latitudes and comparison with velocity shear driven processes," *Journal of Geophysical Research*, vol. 95, p. 7799, A6 1990.

[6] H. Hoang, L. B. Clausen, K. Røed, T. A. Bekkeng, E. Trondsen, B. Lybekk, H. Strøm, D. M. Bang-Hauge, A. Pedersen, A. Spicher, and J. I. Moen, "The multi-needle langmuir probe system on board NorSat-1," *Space Science Reviews*, vol. 214, p. 75, 4 Jun. 2018.

[7] F. F. Chen, "Langmuir probe diagnostics," *Mini-Course on Plasma Diagnostics, IEEE-ICOPS meeting, Jeju, Korea*, 2003.

[8] H. M. Mott-Smith and I. Langmuir, "The theory of collectors in gaseous discharges," *Physical Review*, vol. 28, pp. 727–763, 4 1926.

[9] J. E. Allen, "Probe theory - the orbital motion approach," *Physica Scripta*, vol. 45, pp. 497–503, 5 May 1992.

[10] A. Barjatya, C. M. Swenson, D. C. Thompson, and K. H. Wright, "Invited article: Data analysis of the floating potential measurement unit aboard the international space station," *Review of Scientific Instruments*, vol. 80, 4 2009.

[11] I. D. Sudit and R. C. Woods, "A study of the accuracy of various langmuir probe theories," *Journal of Applied Physics*, vol. 76, pp. 4488–4498, 8 1994.

[12] A. R. Hoskinson and N. Hershkowitz, "Effect of finite length on the current-voltage characteristic of a cylindrical langmuir probe in a multidipole plasma chamber," *Plasma Sources Science and Technology*, vol. 15, pp. 85–90, 1 2006.

[13] A. Barjatya and W. Merritt, "Error analysis of multi-needle langmuir probe measurement technique," *Review of Scientific Instruments*, vol. 89, pp. 1–5, 4 2018.

[14] S. Marholm, R. Marchand, D. Darian, W. J. Miloch, and M. Mortensen, "Impact of miniaturized fixed-bias multineedle langmuir probes on CubeSats," *IEEE Transactions on Plasma Science*, vol. 47, pp. 3658–3666, 8 Aug. 2019.

[15] S. Marholm and R. Marchand, "Finite-length effects on cylindrical langmuir probes," *Physical Review Research*, vol. 2, p. 023 016, 2 Apr. 2020.

[16] R. E. Ergun, L. A. Andersson, C. M. Fowler, and S. A. Thaller, "Kinetic modeling of langmuir probes in space and application to the MAVEN langmuir probe and waves instrument," *Journal of Geophysical Research: Space Physics*, vol. 126, 3 Mar. 2021.

[17] J. G. Laframboise and J. Rubinstein, "Theory of a cylindrical probe in a collisionless magnetoplasma," *Physics of Fluids*, vol. 19, pp. 1900–1908, 12 1976.

[18] J. Rubinstein and J. G. Laframboise, "Theory of a spherical probe in a collisionless magnetoplasma," *Physics of Fluids*, vol. 25, pp. 1174–1182, 7 1982.

[19] R. Marchand, "PTetra, a tool to simulate low orbit satellite–plasma interaction," *IEEE Transactions on Plasma Science*, vol. 40, pp. 217–229, 2 Feb. 2012.

[20] R. Marchand and P. A. R. Lira, "Kinetic simulation of spacecraft–environment interaction," *IEEE Transactions on Plasma Science*, vol. 45, pp. 535–554, 4 Apr. 2017.

[21] P. Frey and P. L. George, *Mesh Generation: Application to Finite Elements*. ISTE, 2007.

[22] C. Geuzaine and J.-F. Remacle, "Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities," *International Journal for Numerical Methods in Engineering*, vol. 79, pp. 1309–1331, 11 Sep. 2009.

[23] D. Bilitza, D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. Reinisch, "The international reference ionosphere 2012 – a model of international collaboration," *Journal of Space Weather and Space Climate*, vol. 4, A07, Feb. 2014.

[24] H. Qin, S. Zhang, J. Xiao, J. Liu, Y. Sun, and W. M. Tang, "Why is boris algorithm so good?" *Physics of Plasmas*, vol. 20, p. 084 503, 8 Aug. 2013.

[25] A. Gron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st. O'Reilly Media, Inc., 2017.

[26] I Goodfellow, Y Bengio, and A Courville, *Deep Learning*. MIT Press, 2016.

[27] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," 2016, pp. 265–283.

[28] R. Cavoretto, T. Schneider, and P. Zulian, "Opencl based parallel algorithm for RBF-PUM interpolation," *Journal of Scientific Computing*, vol. 74, pp. 267–289, 1 Jan. 2018.

[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. USA: Cambridge University Press, 2007.

[30] J. Chalaturnyk and R. Marchand, "Regression-based interpretation of langmuir probe measurements," *Frontiers in Physics*, vol. 7, APR May 2019.

[31] J. Guthrie, R. Marchand, and S. Marholm, "Inference of plasma parameters from fixed-bias multi-needle langmuir probes (m-NLP)," *Measurement Science and Technology*, vol. 32, 9 2021.

[32] A. Olowookere and R. Marchand, "Fixed bias probe measurement of a satellite floating potential," *IEEE Transactions on Plasma Science*, vol. 49, pp. 862–870, 2 2021.

[33] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, pp. 119–139, 6 1951.

[34] R. L. Hardy, "Multiquadric equations of topography and other irregular surfaces," *Journal of Geophysical Research*, vol. 76, pp. 1905–1915, 8 Mar. 1971.

[35] M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review," *Algorithms for approximation*, 1987.

[36] D Broomhead and D Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.

[37] S. Chen, C. Cowan, and P. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, pp. 302–309, 2 Mar. 1991.

[38] L Xu, A Krzyzak, and E Oja, "Rival penalized competitive learning for clustering analysis, rbf net, and curve detection," *IEEE Transactions on Neural Networks*, vol. 4, pp. 636–649, 4 1993.

[39] B. A. Whitehead and T. D. Choate, "Evolving space-filling curves to distribute radial basis functions over an input space," *IEEE Transactions on Neural Networks*, vol. 5, pp. 15–23, 1 1994.

[40] C. Fraley, "Algorithms for model-based gaussian hierarchical clustering," *SIAM Journal on Scientific Computing*, vol. 20, pp. 270–281, 1 Jan. 1998.

[41] G. Celeux and G. Govaert, "Gaussian parsimonious clustering models," *Pattern Recognition*, vol. 28, pp. 781–793, 5 May 1995.

[42] E. A. Lim and Z. Zainuddin, "An improved fast training algorithm for rbf networks using symmetry-based fuzzy c-means clustering," *Matematika*, vol. 24, pp. 141–148, 2 2008.

[43] R. Cavoretto, "Adaptive radial basis function partition of unity interpolation: A bivariate algorithm for unstructured data," *Journal of Scientific Computing*, vol. 87, 2 May 2021.

[44] M. D. Buhmann, *Radial Basis Functions*. Cambridge University Press, Jul. 2003, vol. 12.

[45] G. Fasshauer and M. McCourt, *Kernel-based Approximation Methods using MATLAB*. WORLD SCIENTIFIC, Sep. 2015, vol. 19.

[46] C. S. K. Dash, A. K. Behera, S. Dehuri, and S.-B. Cho, "Radial basis function neural networks: A topical state-of-the-art survey," *Open Computer Science*, vol. 6, pp. 33–63, 1 May 2016.

[47] G. S. Bhatia and G. Arora, "Radial basis function methods for solving partial differential equations-a review," *Indian Journal of Science and Technology*, vol. 9, 45 2016.

[48] L. Clarke, I. Glendinning, and R. Hempel, "The MPI message passing interface standard," K. M. Decker and R. M. Rehmann, Eds., Birkhäuser Basel, 1994, pp. 213–218.

[49] T. Hengl and H. Reuter, *Geomorphometry. concepts, software, applications*, Jun. 2008.

[50] G. Liu, *Segmented flow meter data*, Jan. 2021.

[51] R. Pirjola, "Geomagnetically induced currents during magnetic storms," *IEEE Transactions on Plasma Science*, vol. 28, pp. 1867–1873, 6 2000.

[52] D. Baker, "The occurrence of operational anomalies in spacecraft and their relationship to space weather," *IEEE Transactions on Plasma Science*, vol. 28, pp. 2007–2016, 6 2000.

[53] G. Rudenko and A. Uralov, "Calculation of ionospheric effects due to acoustic radiation from an underground nuclear explosion," *Journal of Atmospheric and Terrestrial Physics*, vol. 57, pp. 225–236, 3 Mar. 1995.

[54] J. B. Davies and C. B. Archambeau, "Modeling of atmospheric and ionospheric disturbances from shallow seismic sources," *Physics of the Earth and Planetary Interiors*, vol. 105, pp. 183–199, 3-4 Jan. 1998.

[55] V. Krasnov and Y. Drobzheva, "The acoustic field in the ionosphere caused by an underground nuclear explosion," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 67, pp. 913–920, 10 Jul. 2005.

[56] M. Parrot, J. Berthelier, J. Lebreton, J. Sauvaud, O. Santolik, and J. Blecki, "Examples of unusual ionospheric observations made by the DEMETER satellite over seismic regions," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 31, pp. 486–495, 4-9 Jan. 2006.

[57] M. Parrot, "Statistical analysis of automatically detected ion density variations recorded by DEMETER and their relation to seismic activity," *Annals of Geophysics*, vol. 55, pp. 149–155, 1 Apr. 2012.

[58] Y.-M. Yang, J. L. Garrison, and S.-C. Lee, "Ionospheric disturbances observed coincident with the 2006 and 2009 north korean underground nuclear tests," *Geophysical Research Letters*, vol. 39, n/a–n/a, 2 Jan. 2012.

[59] K. Ryu, E. Lee, J. S. Chae, M. Parrot, and S. Pulinets, "Seismo-ionospheric coupling appearing as equatorial electron density enhancements observed via DEMETER electron density measurements," *Journal of Geophysical Research: Space Physics*, vol. 119, pp. 8524–8542, 10 Oct. 2014.

[60] X. H. Shen, X. M. Zhang, S. G. Yuan, L. W. Wang, J. B. Cao, J. P. Huang, X. H. Zhu, P. Piergiorgio, and J. P. Dai, "The state-of-the-art of the china seismo-electromagnetic satellite mission," *Science China Technological Sciences*, vol. 61, pp. 634–642, 5 May 2018.

[61] A. D. Santis, D. Marchetti, F. J. Pavón-Carrasco, G. Cianchini, L. Perrone, C. Abbattista, L. Alfonsi, L. Amoruso, S. A. Campuzano, M. Carbone, C. Cesaroni, G. D. Franceschi, A. D. Santis, R. D. Giovambattista, A. Ippolito, A. Piscini, D. Sabbagh, M. Soldani, F. Santoro, L. Spogli, and R. Haagmans, "Precursory worldwide signatures of earthquake occurrences on Swarm satellite data," *Scientific Reports*, vol. 9, p. 20 287, 1 Dec. 2019.

[62] A. J. Hundhausen, J. R. Asbridge, S. J. Bame, H. E. Gilbert, and I. B. Strong, "VELA 3 satellite observations of solar wind ions: A preliminary report," *Journal of Geophysical Research*, vol. 72, p. 87, 1 1967.

[63] Y. I. Galperin, V. N. Ponomarev, and A. G. Zosimova, *Direct measurements of ion drift velocity in the upper ionosphere during a magnetic storm. part 1: Experiment description and some results of measurements during magnetically quiet time*, Oct. 1973.

[64] W. B. Hanson, D. R. Zuccaro, C. R. Lippincott, and S. Sanatani, "The retarding-potential analyzer on atmosphere explorer," *Radio Science*, vol. 8, pp. 333–339, 4 Apr. 1973.

[65] Y. Galperin, V. N. Ponomarev, and A. G. Zosimova, "Plasma convection in polar ionosphere," *Annales de Geophysique*, vol. 30, pp. 1–7, 1 1974.

[66] W. B. Hanson and R. A. Heelis, "Techniques for measuring bulk gas-motions from satellites," *Space Science Instrumentation*, vol. 1, pp. 493–524, Nov. 1975.

[67] R. A. Heelis, W. B. Hanson, C. R. Lippincott, D. R. Zuccaro, L. H. Harmon, B. J. Holt, J. E. Doherty, and R. A. Power, "The ion drift meter for dynamics Explorer-B.," *Space Science Instrumentation*, vol. 5, pp. 511–521, Dec. 1981.

[68] K. W. Ogilvie, D. J. Chornay, R. J. Fritzenreiter, F. Hunsaker, J. Keller, J. Lobell, G. Miller, J. D. Scudder, E. C. Sittler, R. B. Torbert, D. Bodet, G. Needell, A. J. Lazarus, J. T. Steinberg, J. H. Tappan, A. Mavretic, and E. Gergin, "SWE, a comprehensive plasma instrument for the WIND spacecraft," *Space Science Reviews*, vol. 71, pp. 55–77, 1-4 Feb. 1995.

[69] C. Reigber, H. Lühr, and P. Schwintzer, Eds., *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*. Springer Berlin Heidelberg, 2003, pp. 212–219.

[70] J. Berthelier, M. Godefroy, F. Leblanc, E. Seran, D. Peschard, P. Gilbert, and J. Artru, "IAP, the thermal plasma analyzer on DEMETER," *Planetary and Space Science*, vol. 54, pp. 487–501, 5 Apr. 2006.

[71] R. A. Stoneback, R. L. Davidson, and R. A. Heelis, "Ion drift meter calibration and photoemission correction for the C/NOFS satellite," *Journal of Geophysical Research: Space Physics*, vol. 117, n/a–n/a, A8 Aug. 2012.

[72] R. A. Heelis, R. A. Stoneback, M. D. Perdue, M. D. Depew, W. A. Morgan, M. W. Mankey, C. R. Lippincott, L. L. Harmon, and B. J. Holt, "Ion velocity measurements for the ionospheric connections explorer," *Space Science Reviews*, vol. 212, pp. 615–629, 1-2 Oct. 2017.

[73] R. A. Heelis and W. B. Hanson, *Measurements of thermal ion drift velocity and temperature using planar sensors*, Mar. 2013.

[74] M. Satir, F. Sik, E. Turkoz, and M. Celik, "Design of the retarding potential analyzer to be used with BURFIT-80 ion thruster and validation using PIC-DSMC code," IEEE, Jun. 2015, pp. 577–582.

[75] E. Séran, "Reconstruction of the ion plasma parameters from the current measurements: Mathematical tool," *Annales Geophysicae*, vol. 21, pp. 1159–1166, 5 2003.

[76] C. Carlson, D. Curtis, G. Paschmann, and W. Michel, "An instrument for rapidly measuring plasma distribution functions with high resolution," *Advances in Space Research*, vol. 2, pp. 67–70, 7 Jan. 1982.

[77] C. W. Carlson, J. P. Mcfadden, P. Turin, D. W. Curtis, and A. Magoncelli, "The electron and ion plasma experiment for FAST," *Space Science Reviews*, vol. 98, pp. 33–66, 1-2 2001.

[78] C. W. Carlson and J. P. McFadden, *Design and application of imaging plasma instruments*, Mar. 2013.

[79] B. A. Whalen, D. J. Knudsen, A. W. Yau, A. M. Pilon, T. A. Cameron, J. F. Sebesta, D. J. McEwen, J. A. Koehler, N. D. Lloyd, G. Pocobelli, J. G. Laframboise, W. Li, R. Lundin, L. Eliasson, S. Watanabe, and G. S. Campbell, "The Freja F3C cold plasma analyzer," *Space Science Reviews*, vol. 70, pp. 541–561, 3-4 Nov. 1994.

[80] A. W. Yau, G. P. Garbe, M. J. Greffen, W. Miyake, and E. Sagawa, "Planned observations of thermal plasma drifts and solar wind interactions in the martian ionosphere," *Earth, Planets and Space*, vol. 50, pp. 195–198, 3 Mar. 1998.

[81] D. J. Knudsen, J. K. Burchill, K. Berg, T. Cameron, G. A. Enno, C. G. Marcellus, E. P. King, I. Wevers, and R. A. King, "A low-energy charged particle distribution imager with a compact sensor for space applications," *Review of Scientific Instruments*, vol. 74, pp. 202–211, 1 Jan. 2003.

[82] A. W. Yau, A. Howarth, A. White, G. Enno, and P. Amerl, "Imaging and rapid-scanning ion mass spectrometer (IRM) for the CASSIOPE e-POP mission," *Space Science Reviews*, vol. 189, pp. 41–63, 1-4 2015.

[83] D. J. Knudsen, J. K. Burchill, S. C. Buchert, A. I. Eriksson, R. Gill, J. Wahlund, L. Åhlen, M. Smith, and B. Moffat, "Thermal ion imagers and langmuir probes in the Swarm electric field instruments," *Journal of Geophysical Research: Space Physics*, vol. 122, pp. 2655–2673, 2 Feb. 2017.

[84] E. Séran, J.-J. Berthelier, F. Z. Saouri, and J.-P. Lebreton, "The spherical segmented langmuir probe in a flowing thermal plasma: Numerical model of the current collection," *Annales Geophysicae*, vol. 23, pp. 1723–1733, 5 Jul. 2005.

[85] J.-P. Lebreton, S. Stverak, P. Travnicek, M. Maksimovic, D. Klinge, S. Merikallio, D. Lagoutte, B. Poirier, P.-L. Blelly, Z. Kozacek, and M. Salaquarda, "The ISL langmuir probe experiment processing onboard DEMETER: Scientific objectives, description and first results," *Planetary and Space Science*, vol. 54, pp. 472–486, 5 Apr. 2006.

[86] S. Santandrea, K. Gantois, K. Strauch, F. Teston, E. Tilmans, C. Baijot, D. Gerrits, A. D. Groof, G. Schwehm, and J. Zender, "PROBA2: Mission and spacecraft overview," *Solar Physics*, vol. 286, pp. 5–19, 1 Aug. 2013.

[87] J. Berthelier, M. Godefroy, F. Leblanc, M. Malingre, M. Menvielle, D. Lagoutte, J. Brochot, F. Colin, F. Elie, C. Legendre, P. Zamora, D. Benoist, Y. Chapuis, J. Artru, and R. Pfaff, "ICE, the electric field experiment on DEMETER," *Planetary and Space Science*, vol. 54, pp. 456–471, 5 Apr. 2006.

[88] R. E. Ergun, C. W. Carlson, F. S. Mozer, G. T. Delory, M. Temerin, J. P. Mcfadden, D. Pankow, R. Abiad, P. Harvey, R. Wilkes, H. Primbsch, R. Elphic, R. Strangeway, R. Pfaff, and C. A. Cattell, "The FAST satellite fields instrument," *Space Science Reviews*, vol. 98, pp. 67–91, 1-2 2001.

[89] A. I. Eriksson, M. André, B. Klecker, H. Laakso, P.-A. Lindqvist, F. Mozer, G. Paschmann, A. Pedersen, J. Quinn, R. Torbert, K. Torkar, and H. Vaith, "Electric field measurements on cluster: Comparing the double-probe and electron drift techniques," *Annales Geophysicae*, vol. 24, pp. 275–289, 1 Mar. 2006.

[90] C. R. Englert, D. D. Babcock, and J. M. Harlander, "Doppler asymmetric spatial heterodyne spectroscopy (DASH): Concept and experimental demonstration," *Applied Optics*, vol. 46, p. 7297, 29 Oct. 2007.

[91] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second. Society for Industrial and Applied Mathematics, Jan. 2003.

[92] S. Lai and M. Tautz, "Aspects of spacecraft charging in sunlight," *IEEE Transactions on Plasma Science*, vol. 34, pp. 2053–2061, 5 Oct. 2006.

[93] K. Torkar, R. Nakamura, M. Tajmar, C. Scharlemann, H. Jeszenszky, G. Laky, G. Fremuth, C. P. Escoubet, and K. Svenes, "Active spacecraft potential control investigation," *Space Science Reviews*, vol. 199, pp. 515–544, 1-4 Mar. 2016.

[94] G. D. Reeves, G. L. Delzanno, P. A. Fernandes, K. Yakymenko, B. E. Carlsten, J. W. Lewellen, M. A. Holloway, D. C. Nguyen, R. F. Pfaff, W. M. Farrell, D. E. Rowland, M. Samara, E. R. Sanchez, E. Spanswick, E. F. Donovan, and V. Roytershteyn, "The beam plasma interactions experiment: An active experiment using pulsed electron beams," *Frontiers in Astronomy and Space Sciences*, vol. 7, p. 23, Jun. 2020.

[95] D. B. Beard and F. S. Johnson, "Ionospheric limitations on attainable satellite potential," *Journal of Geophysical Research*, vol. 66, pp. 4113–4122, 12 Dec. 1961.

[96] P. A. R. Lira and R. Marchand, "Simulation inference of plasma parameters from langmuir probe measurements," *Earth and Space Science*, vol. 8, 3 Mar. 2021.

[97] N. A. Barkhatov and S. E. Revunov, "Neural network classification of discontinuities in space plasma parameters," *Geomagnetism and Aeronomy*, vol. 50, pp. 894–904, 7 Dec. 2010.

[98] H. Breuillard, R. Dupuis, A. Retino, O. L. Contel, J. Amaya, and G. Lapenta, "Automatic classification of plasma regions in near-earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observations," *Frontiers in Astronomy and Space Sciences*, vol. 7, Sep. 2020.

[99] S. Bhattarai and L. N. Mishra, "Theoretical study of spherical langmuir probe in maxwellian plasma," *International Journal of Physics*, vol. 5, pp. 73–81, 3 Aug. 2017.

[100] P. A. R. Lira, R. Marchand, J. Burchill, and M. Forster, "Determination of Swarm front plate's effective cross section from kinetic simulations," *IEEE Transactions on Plasma Science*, vol. 47, pp. 3667–3672, 8 2019.

[101] J. D. Johnson and A. J. T. Holmes, "Edge effect correction for small planar langmuir probes," *Review of Scientific Instruments*, vol. 61, pp. 2628–2631, 10 Oct. 1990.

[102] T. E. Sheridan, "The plasma sheath around large discs and ion collection by planar langmuir probes," *Journal of Physics D: Applied Physics*, vol. 43, p. 105 204, 10 Mar. 2010.

[103] D. L. Rudakov, J. A. Boedo, R. A. Moyer, R. D. Lehmer, G. Gunner, and J. G. Watkins, "Fast electron temperature diagnostic based on langmuir probe current harmonic detection on DIII-D," *Review of Scientific Instruments*, vol. 72, pp. 453–456, 1 II 2001.

[104] J. E. Allen, R. L. F. Boyd, and P Reynolds, "The collection of positive ions by a probe immersed in a plasma," *Proceedings of the Physical Society. Section B*, vol. 70, pp. 297–304, 3 Mar. 1957.

[105] F. F. Chen, "Numerical computations for ion probe characteristics in a collisionless plasma," *Journal of Nuclear Energy. Part C, Plasma Physics, Accelerators, Thermonuclear Research*, vol. 7, pp. 47–67, 1 Jan. 1965.

[106] F. F. Chen, "Lecture notes on langmuir probe diagnostics," Mini-Course on Plasma Diagnostics, IEEEICOPS meeting, 2003.

[107] I. B. Bernstein and I. N. Rabinowitz, "Theory of electrostatic probes in a low-density plasma," *Physics of Fluids*, vol. 2, p. 112, 2 1959.

[108] J. G. Laframboise, "Theory of spherical and cylindrical langmuir probes in a collisionless, maxwellian plasma at rest," *Toronto Univ. (Ontario). Inst. for Aerospace Studies*, 1966.

[109] T. A. Bekkeng, K. S. Jacobsen, J. K. Bekkeng, A. Pedersen, T. Lindem, J. P. Lebreton, and J. I. Moen, "Design of a multi-needle langmuir probe system," *Measurement Science and Technology*, vol. 21, p. 085 903, 8 Aug. 2010.

[110] A. Barjatya, J. P. St-Maurice, and C. M. Swenson, "Elevated electron temperatures around twin sporadic e layers at low latitude: Observations and the case for a plausible link to currents parallel to the geomagnetic field," *Journal of Geophysical Research: Space Physics*, vol. 118, pp. 7316–7328, 11 2013.

[111] S. Debchoudhury, A. Barjatya, J. I. Minow, V. N. Coffey, and M. O. Chandler, "Observations and validation of plasma density, temperature, and o+ abundance from a langmuir probe onboard the international space station," *Journal of Geophysical Research: Space Physics*, vol. 126, 10 Oct. 2021.

[112] G. Liu and R. Marchand, "Kinetic simulation of segmented plasma flow meter response in the ionospheric plasma," *Journal of Geophysical Research: Space Physics*, vol. 126, 5 2021.

[113] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Y. Bengio and Y. LeCun, Eds., 2015.

[114] M. F. Ivarsen, H. Hoang, L. Yang, L. B. Clausen, A. Spicher, Y. Jin, E. Trondsen, J. I. Moen, K. Roed, and B. Lybekk, "Multineedle langmuir probe operation and acute probe current susceptibility to spacecraft potential," *IEEE Transactions on Plasma Science*, vol. 47, pp. 3816–3823, 8 2019.

Appendix A: CERBF Implementations

This appendix describes the implementation of Message Passing Interface (MPI) in the CERBF program. The CERBF program consists of several subprograms that are called in sequence in the main program to perform the training and validation procedures. The purposes of the essential subprograms are discussed in the order in which they are called in the main program as follows:

At the beginning of the program, the control parameters in the 'rbf.dat' file are read in the main program, then the subprogram **readtrainingdata** is called.

readtrainingdata The purpose of this subprogram is to read and prepare the data for training. The subprogram will read the training data only once at the beginning of the program. The main processor will select a random subset from the training set and broadcast the indexes of the selected elements to all processors each time this function is called. After the first iteration, the randomly selected subset has to be distinguished from the optimal centers carried from the previous iteration to avoid identical elements in the list. The first N nodes in the selected subset will be replaced by the optimal centers from the previous iteration.

The CERBF program uses multiple processors to share the workload by distributing the combination lists to the different processors. This is achieved using the **combinationStep** subprogram.

combinationStep This subprogram will move the list containing the indexes of the current combination forward once. For example, if the training data contains 5 nodes, and 3 centers are used to approximate a function, then the first list of 5 choose 3 is (1, 2, 3). After applying this subprogram once, the list will become (1, 2, 4), and after applying it twice, the list will become (1, 2, 5), and so on.

To initialize a training, every processor will contain the list of (1, 2, 3). Then, the nth processor will run the **combinationStep** subprogram $n-1$ times. For example, if there are three processors, then the second processor will run this subprogram once and have the list (1, 2, 4). The third processor will run this subprogram twice to have the list (1, 2, 5). Then subprogram **stepcal** will be called.

stepcal For the current list, the matrix of Eq. 2.15 is solved, and the cost functions are calculated. The lowest cost function and the list that this cost function is calculated are saved in each processor.

In the following steps, each processor will run the **combinationStep** subprogram the total number of the processors times and then call **stepcal** subprogram, till all possible combinations have been gone through. For example, following the previous

example, in processor one, run the **combinationStep** subprogram three times, and the list in processor one becomes (1, 3, 4). The lists in the second and third processors become (1, 3, 5), and (1, 4, 5). This way, the work to go through all possible combinations is split between different processors. The main program checks if all combinations have been gone through after each one hundred times that the **stepcal** subprogram is called to minimize the number of if statement in the program. After all combinations have been gone through, the **gradientdescent** subprogram is performed to further optimize the model in each processor.

gradientdescent This subprogram accounts for the "nugget effect". The cost function is minimized using the gradient descent method.

Then the subprogram **gathertoroot** will be called to gather the necessary information from all processors to the root processor (1st processor). Then in the main program, the root processor will determine if the lowest cost function from the current CERBF iteration is better than previous iterations and save it if it is. The $\bar{X}$ and Y of the optimal centers are also broadcast to all processors. The root processor will then determine if the training process is ended. If not, the program will select another subset using the **readtrainingdata** subprogram and start over a new CERBF iteration. If the maximum number of training iterations is reached, then the program will do validation or inference if the corresponding files are provided using the **validateorinference** subprogram. The program is a little over one thousand lines. A schematic of CERBF is shown in Fig. 3.3. The complete program, with some sample data and results, can be found at https://codeocean.com/capsule/7662817/tree.