m-NLP inference models using simulation and regression techniques

Guangdong Liu[1,1,1], Sigvald Marholm[2,2,2], Anders Eklund[3,3,3], Lasse Boy Novock Clausen[4,4,4], and Richard Marchand[1,1,1]

[1]University of Alberta
[2]University Of Oslo
[3]SINTEF Industry
[4]University of Oslo

November 30, 2022

Abstract

Current inference techniques for processing multi-needle Langmuir Probe (m-NLP) data are often based on 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 synthetic data sets, which are then used to train and validate regression-based models capable of inferring electron density and satellite potentials from 4-tuples of currents collected with fixed-bias needle probes similar to those on the NorSat-1 satellite. Based on our synthetic data, the techniques presented enable excellent inferences of the plasma density, and floating potentials, while the generally accepted OML inferred densities are approximately three times too high. The new inference techniques that we propose, are applied to NorSat-1 data, and compared with OML inferences. While both regression and OML based inferences of floating potentials agree well with synthetic data, only regression inferred potentials are consistent with satellite measured currents, indicating that the regression based inference models are more robust and accurate when applied to satellite data.

m-NLP inference models using simulation and regression techniques

Guangdong Liu[1], Sigvald Marholm [2,3], Anders J. Eklund[4], Lasse Clausen[2], Richard Marchand[1]

[1]Department of Physics, University of Alberta, Edmonton, AB, Canada
[2]Department of Physics, University of Oslo, Oslo, Norway
[3]Department of Computational Materials Processing, Institute for Energy Technology, Kjeller, Norway
[4]Materials Physics Oslo, SINTEF Industry, Oslo, Norway

Key Points:

- 3-D kinetic PIC simulations are used to simulate currents collected by m-NLP in order to create a synthetic solution library
- Models to infer physical parameters from m-NLP measurements are constructed and assessed on the basis of synthetic and in situ data sets
- Promising new approaches are identified to analyze m-NLP measurements based on simulation and in-situ data

Corresponding author: Guangdong Liu, guangdon@ualberta.ca

Abstract

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 a synthetic data set, regression-based models capable of inferring electron density and satellite potentials 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 show promising results for plasma density inferences with RMS relative errors less than 20 %, and satellite potential inferences with RMS errors less than 0.2 V for potentials ranging from -6 V to -1 V. The new inference approaches presented are applied to NorSat-1 data, and compared with existing state-of-the-art inference techniques.

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 (Bhattarai & Mishra, 2017), cylindrical (Hoang, Clausen, et al., 2018), and planar probes (Lira et al., 2019; Johnson & Holmes, 1990; Sheridan, 2010). Probes can be operated in sweep mode (Lebreton et al., 2006), harmonic mode (Rudakov et al., 2001), or fixed biased mode (Jacobsen et al., 2010), for different types of missions and measurements. 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) (Mott-Smith & Langmuir, 1926) theory, the Allen-Boyd-Reynolds (ABR) theory (Allen et al., 1957; Chen, 1965, 2003), and the Bernstein-Rabinowitz-Laframboise (BRL) theory (Bernstein & Rabinowitz, 1959; Laframboise, 1966) 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 (Hoang, Røed, et al., 2018; Jacobsen et al., 2010). 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 (Jacobsen et al., 2010). 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:

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

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 (Marholm & Marchand, 2020; Hoang, Røed, et al., 2018). 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 (Jacobsen et al., 2010). 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 (Bekkeng et al., 2010; Hoang, Clausen, et al., 2018; Hoang et al., 2019), 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 (Barjatya et al., 2009). 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. 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 Deybe 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 (Sudit & Woods, 1994). 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. 1 using a β value of 0.67 instead of 0.55 calculated in a stationary plasma (Ergun et al., 2021). 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 (Hoang, Røed, et al., 2018). Simulation results by Marholm et al. showed that even a 50 mm probe at rest can be characterized by a $\beta \sim 0.8$ (Marholm et al., 2019), 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 they are 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 (Marholm & Marchand, 2020). 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 (Hoskinson & Hershkowitz, 2006). 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 (Barjatya & Merritt, 2018). In what follows, we find that densities estimated using Eq. 1 assuming $\beta = 0.5$ are about three times larger than the known values used as input in our simulations, as illustrated in section 3.1. This is consistent with findings in (Barjatya & Merritt, 2018; Guthrie et al., 2021), 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 led to remarkable agreement with density measured using a radio frequency impedance probe on the international space station (Barjatya et al., 2009, 2013; Debchoudhury et al., 2021). 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 (Barjatya et al., 2009; Barjatya & Merritt, 2018; Hoang, Røed, et al., 2018).

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 con-

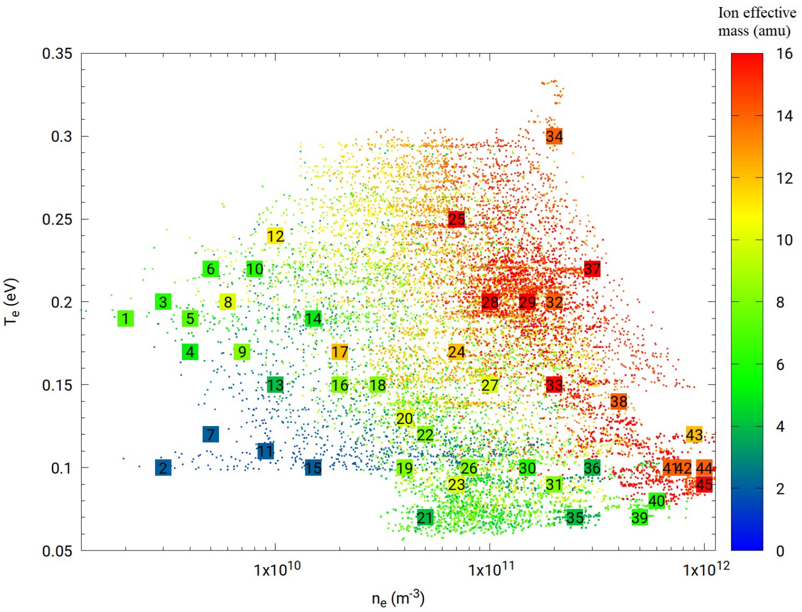

Figure 1. Scatter plot of plasma parameters obtained from the IRI model, corresponding to different latitudes, longitudes, altitudes, and times, as listed in Table 1. The x and y axes, and the color bar refer respectively, to the electron density, electron temperature, and the ion effective mass. Numbered squares identify the set of parameters used in the kinetic simulations.

struction of a synthetic data set are presented in Sec. 2. In Sec. 3, regression models are trained using synthetic data sets, and they are assessed using distinct validation sets. In Sec. 4, the same models are applied to NorSat-1 data, to infer densities and satellite potentials from in situ measured currents. Section 5 summarizes our findings and presents some concluding remarks.

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. We then describe the various models to infer density, and satellite potential from m-NLP measurement.

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) (Bilitza et al., 2014) model in a broad range of latitudes, longitudes, altitudes, and times as shown in Fig. 1. The ranges considered for these parameters are summarized in Tab. 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. 1. The three-dimensional PIC code PTetra (Marchand, 2012; Marchand & Lira, 2017) is used to simulate probe currents in this study. Cross comparisons are made between PTetra simulation results and analytic results under conditions when those are valid, and with other independently developed simulation codes, and show excellent agreement (Deca et al., 2013; Marchand et al., 2014).

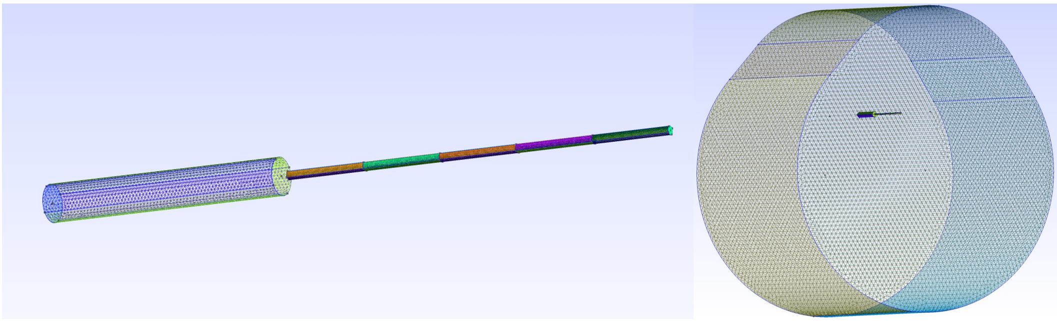

Figure 2. Illustration of a m-NLP geometry (left), and the simulation domain (right). The needle probe has a length of 25 mm and a radius of 0.255 mm, with a guard of 15 mm in length and 1.1 mm in radius. The ram flow is from the top of the simulation domain and is assumed to be 7500 m/s.

In PTetra, space is discretized using unstructured adaptive tetrahedral meshes (Frey & George, 2007; Geuzaine & Remacle, 2009). Poisson's equation is solved at each time step using Saad's GMRES sparse matrix solver (Saad, 2003) 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. 2, is generated with GMSH (Geuzaine & Remacle, 2009). 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. 2. The simulations are made using two different domain sizes depending on the Debye length of the plasma. For plasma densities below 2×10^{10} m^{-3} corresponding to a Debye length of 1.9-7.2 cm, a lager domain is used. For plasma densities 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. 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 a simulation, due to particles being collected, leaving, or entering the domain. In the simulations, the probe is segmented into five segments of equal lengths, 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. The probes are assumed to be sufficient far on the ram side, away from other satellite components, to be unaffected by their presence. NorSat-1 satellite has a Sun synchronous orbit, thus is moving approximately parallel to the magnetic field near the equator. As a result, in these regions $\vec{V}\times\vec{B}$ should be small at low and mid magnetic latitudes, and it is not accounted for in the simulations.

Table 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	$-90°$ - $+90°$ 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 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

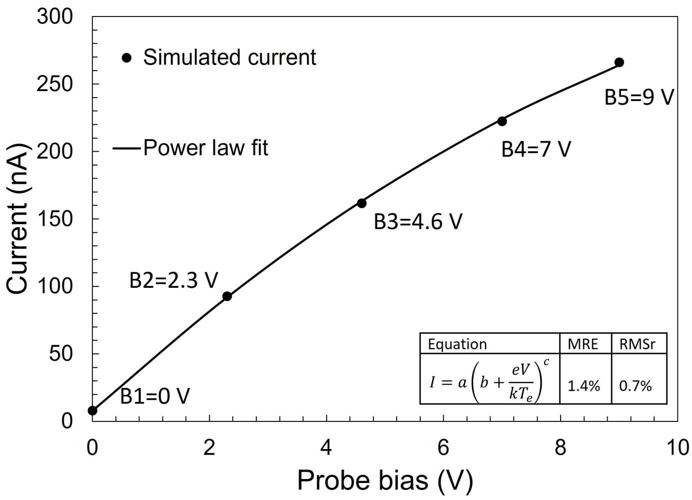

Figure 3. Comparison between calculated currents from PIC simulations, and fitted values using Eq. 6, assuming a density of 2 $\times$ 10^{10} m^{-3}, an effective mass of 8 amu, an electron and ion temperatures of 0.15 and 0.12 eV respectively, corresponding to point 16 in Fig. 1. The fitting errors in the figure are calculated over all 45 sets of plasma conditions using Eq. 3 and 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 non-negative, 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,

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

the root mean square relative error

$$RMSr = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} \frac{(Y_{mod_i} - Y_{data_i})^2}{Y_{mod_i}^2}}, \tag{3}$$

the maximum absolute error

$$MAE = \max \left\{ |Y_{mod} - Y_{data}| \right\}, \tag{4}$$

and the maximum relative error

$$MRE = \max \left\{ \left| \frac{Y_{mod} - Y_{data}}{Y_{mod}} \right| \right\}, \tag{5}$$

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 of the 45 sets of plasma conditions corresponding to squares in Fig. 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:

$$I = a \left(b + \frac{eV}{kT_e} \right)^c, \tag{6}$$

where a, b, and c are adjustable fitting parameters. The MRE calculated for the 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. 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 and the probe bias $V = V_f + V_b$. In simulations, probe currents calculated for voltages with respect to background plasma in the range between 0 to 9 volts are considered as shown in Fig. 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. 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. 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.

2.3 Multivariate regression

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

Table 3. Example entries of the synthetic data set, with currents I_1, I_2, I_3, and I_4 calculated using Eq. 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. 6. The first and second entries correspond respectively to points 16 and 21 in Fig. 1.

$I_1(nA)$	$I_2(nA)$	$I_3(nA)$	$I_4(nA)$	$V_f(V)$	$n_e(\mathrm{m}^{-3})$	$T_e(\mathrm{eV})$	$T_i(\mathrm{eV})$	$\mathrm{m_{eff}}$(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

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 minimizing 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.

2.3.1 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 (Liu & Marchand, 2021; Olowookere & Marchand, 2021; Chalaturnyk & Marchand, 2019; Guthrie et al., 2021). A general expression for RBF regression for a set of independent n-tuples $\bar{X}$ and corresponding dependent variable Y is given by:

$$Y = \sum_{i=1}^{N} a_i G\left(\left|\bar{X} - \bar{X}_i\right|\right). \tag{7}$$

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. 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

$$\sum_{i=1}^{N} a_i G(|\bar{X}_k - \bar{X}_i|) = Y_k \tag{8}$$

for $k = 1,\ \ldots,\ N$. 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

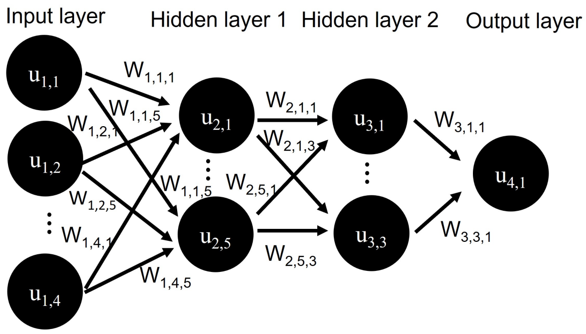

Figure 4. Schematic of a feedforward neural network.

approach, the number of combinations required for $\mathcal{N}$ data points and N centers is given by

$$\binom{\mathcal{N}}{N} = \frac{\mathcal{N}!}{N!(\mathcal{N}-N)!}. \tag{9}$$

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 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 (Liu & Marchand, 2022). 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.

2.3.2 Feedforward neural network

The second multivariate regression approach is a Feedforward neural network as illustrated in Fig. 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

$$u_{i+1,k} = f\left(\sum_{j=1}^{n_i} w_{i,j,k} u_{i,j} + b_{i,k}\right), \tag{10}$$

where $w_{i,j,k}$ are weight factors, $b_{i,j}$ are bias terms, and f is a nonlinear activation function (Goodfellow et al., 2016). 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 (Abadi et al., 2016) with Adam optimizer (Kingma & Ba, 2015), 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.

2.4 Space plasma and satellite parameter inference models

The next step is to construct models that maps the measured currents to the corresponding plasma and satellite conditions in the solution library. Various models used to infer plasma densities and satellite potentials are described in this section.

2.4.1 Density inference

The density can be inferred directly from the above two multivariate regression models using the currents collected by the four probes as inputs. The density can also be inferred using Eq. 1 which can be rewritten as

$$\frac{n_e}{T_e^{\beta-\frac{1}{2}}} = \sqrt{\frac{\pi^2 m_e}{2A^2e^3}} \left(\frac{I_1^{\frac{1}{\beta}} - I_2^{\frac{1}{\beta}}}{V_1 - V_2} \right)^{\beta}. \tag{11}$$

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

$$n_e = \sqrt{\frac{\pi^2 m_e}{2A^2e^3}} \sqrt{\frac{I_1^2 - I_2^2}{V_1 - V_2}}, \tag{12}$$

which gives an expression for the electron density, independently of the temperature (Jacobsen et al., 2010). 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 (Jacobsen et al., 2010). This will be referred as the “Jacobsen linear fit” (JLF) approach. It is now well known however, that for finite length probes, with lengths not much larger than the Debye length, β typically ranges between 0.6 and 1. This is the case in particular for the needle probes on NorSat-1 with ratios between probe lengths to Debye length ranging from 0.5 to 12.5. As a consequence, when this method is applied to the solution library, the inferred density is typically three times larger than the density used in the simulation as shown with red boxes in Fig 5. Analytic inferences can be improve by adopting a boosting strategy. With this approach, the less accurate analytic model is used as a first approximation, which is then corrected by applying a more advanced regression technique.

Two boosting strategies are used in this study, consisting of i) an affine transformation, and b) RBF. Considering that the Pearson correlation coefficient R is invariant under an affine transformation, it follows that the offset between two data set, with a high value of R, can be significantly reduced with a simple affine transformation. To be specific, in this case, the density is first approximated using the JLF approach, and an affine transformation is applied to the natural log of the density as in:

$$\ln(n_e^{\text{affine}}) = a \ln(n_e^{\text{JLF}}) + b. \tag{13}$$

where a and b are determined with a simple least square fit to the known log of the densities in the data set. In the second approach, RBF is used to model the discrepancy between the JLF approximated density and the known densities, and the modeled discrepancy is used to correct the first JLF estimate.

The nonlinear least square fit proposed by Barjatya is also used to infer the density and the satellite potential. In their paper, Barjatya, et al. (Barjatya et al., 2009) apply this method to a full characteristic, covering the ion saturation, the electron retardation, and electron saturation regions. This enabled them to infer all four parameters in Eq. 1, namely, n_e, T_e, V_f, and β. In our analysis, inferences are made from only four currents from four probes at fixed bias voltages, all in the electron saturation region. As shown by Barjatya and Merritt (Barjatya & Merritt, 2018), 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. 11). A solution, proposed in (Barjatya

& Merritt, 2018; Hoang, Røed, et al., 2018), 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 parameters (Barjatya & Merritt, 2018). Thus unless stated otherwise, we assume a fixed electron temperature (~ 2000 K), which is in the middle of the temperature range considered in the simulations, and fit 4-tuples of currents using V_f, n_e, and β values as fitting parameters. This "Barjatya nonlinear fit" (BNLF) approach will be used to infer both density and potential from a set of measurement.

2.4.2 Analytic estimate of V_f

The satellite potential can be inferred directly from the currents using RBF regression. In this approach, the four currents are normalized by dividing every current by their sum, in order to remove the strong density dependence on the currents. A neural network does not produce satisfactory in this case, and it is not used to infer the satellite potential. The floating potential of the spacecraft can also be inferred using the OML equation, by rewriting equation 1 as:

$$V_f \approx V_f + \frac{kT_e}{e} = \frac{V_2 I_1^{\frac{1}{\beta}} - V_1 I_2^{\frac{1}{\beta}}}{I_2^{\frac{1}{\beta}} - I_1^{\frac{1}{\beta}}} = \frac{V_3 I_2^{\frac{1}{\beta}} - V_2 I_3^{\frac{1}{\beta}}}{I_3^{\frac{1}{\beta}} - I_2^{\frac{1}{\beta}}}. \tag{14}$$

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 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. 14 provides a first approximation of V_f (Guthrie et al., 2021). This will be referred to as the "adapted OML" approach.

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.

3.1 Density and satellite potential inference

Direct RBF regression is applied to infer the 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 are 17% and 35%, respectively. A test is made to infer the density using RBF with 35 randomly selected entries from the 45 plasma conditions in the solution library. With 30 voltages, and the same G function, cost function and number of centers, the calculated MRE is also 35 %. This is an indication that 45 sets of plasma conditions and 160 voltages should be sufficient in terms of sampling size to construct regression models. 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 by the largest density. The densities calculated using the synthetic solution library, as well as the cost function are shown in Fig. 5.

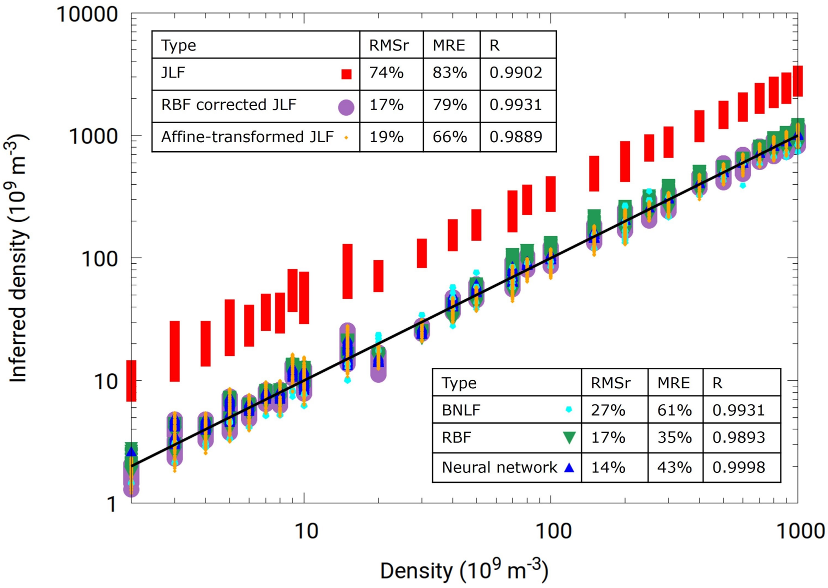

Figure 5. Correlation plot for the density inferences made with different techniques applied to our synthetic validation set. The Pearson correlation coefficient R is calculated using the inferred densities and the density used in the simulation. Black line represent idealized perfect correlation line.

When using an affine transformation to boost the JLF method, the coefficients a and b in Eq. 13 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. When boosting JLF density with RBF, the 4-tuple of currents is used as input variable $\bar{X}$. 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 boosting methods are comparable, but an obvious advantage of using an affine transformation is its simplicity.

The Python 3 `LMFIT` package is used to do the nonlinear fit for the BNLF approaches as in (Debchoudhury et al., 2021). In the fits, the initial values for the density, the potential and the β value are 5×10^{10} m^{-3}, -3 V and 0.85, and the lower and upper bounds are 1×10^9 to 1×10^{12} m^{-3}, -6 to -1 V, and 0.7 to 1, respectively. The potential lower bound of -6 V is needed to ensure that the values under exponent in Eq. 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 for the 3600×4 currents is 0.2 %, and only 1.7% 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 those from the multivariate regression models. The β values calculated are in the range of 0.75 to 1. Using `LMFIT`, approximately 160 sets of currents can be fitted in one 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 (7400 sets can be fitted in one second using an AMD 5800x), 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. Compared to the other density mod-

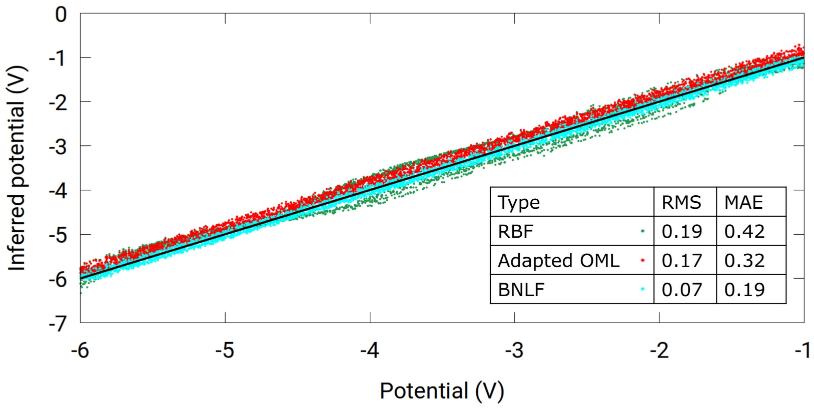

Figure 6. Correlation plot obtained for satellite potential inferred with RBF and OML techniques.

els 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 therefore be of interest in routine data analysis.

When the adapted OML approach is used to infer satellite potentials, a MAE of 0.3 V is calculated using currents collected with probe biases of 10, 9, and 8 volts probes. Referring to Eq. 14, the error of 0.3 V is likely due in part 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. RBF regression is also used to infer satellite potentials. In this case, using $G(x) = |x|$, 5 centers, and minimizing the 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 0.19 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, adapted OML, and BNLF approaches is shown in Fig. 6. All methods show good agreement with values from the synthetic solution library.

3.2 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 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. 2.4. 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. 7, shows back-inferred currents (green) calculated for a probe with 10 V bias against known currents from synthetic data. For com-

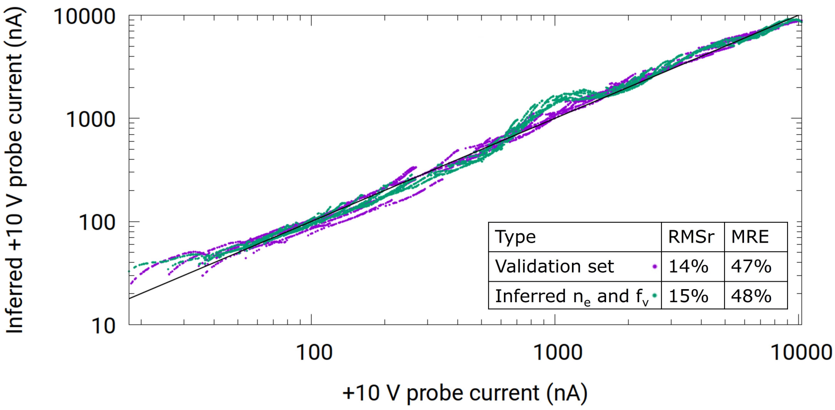

Figure 7. Correlation plot of inferred +10 V probe current against +10 V probe current from the synthetic data set. The calculated +10 probe currents in purple curve is calculated using the validation data set, while the green curve is calculated using inferred densities and floating potentials from RBF regression.

parison, 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 synthetic 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. 7 is seen as confirmation of the validity of our regression models.

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 (Hoang, Clausen, et al., 2018). 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, adapted OML, and the two corrected JLF approaches constructed in 3.1, are also presented.

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. 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 (Hoang, Clausen, et al., 2018). 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. 3.2, and presented below in Sec. 4.3.

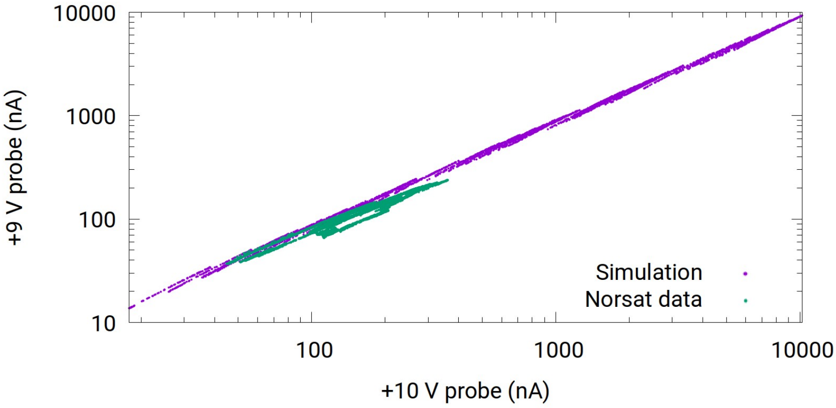

Figure 8. Correlation plot between currents collected by the +9 V and the +10 V probes for both NorSat-1, and synthetic data.

4.2 Density and satellite potential inference

Our models, trained with synthetic data as described in Sec. 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. 3 are shown in Fig. 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.

Applying the BNLF method with only four probes at fixed bias voltages, all in the electron saturation region, is more challenging than applying the technique to a probe operated in sweep mode, covering the ion saturation, the elecron retardation, and the electron saturation regions. The reason is that in sweep mode, characteristics contain much more information than in fixed bias mode, with only four probes. In practice, inferences made from sweep mode characteristics are less sensitive to random errors in the currents, which, owing to their larger numbers, tend to cancel. With only four currents, however, noise is less likely to cancel, and inferences will be more sensitive to errors or noises, in measured currents. For example, the +8 V NorSat-1 probe currents are often slightly lower than expected for a downward concavity in 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. 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. 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 0.41 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

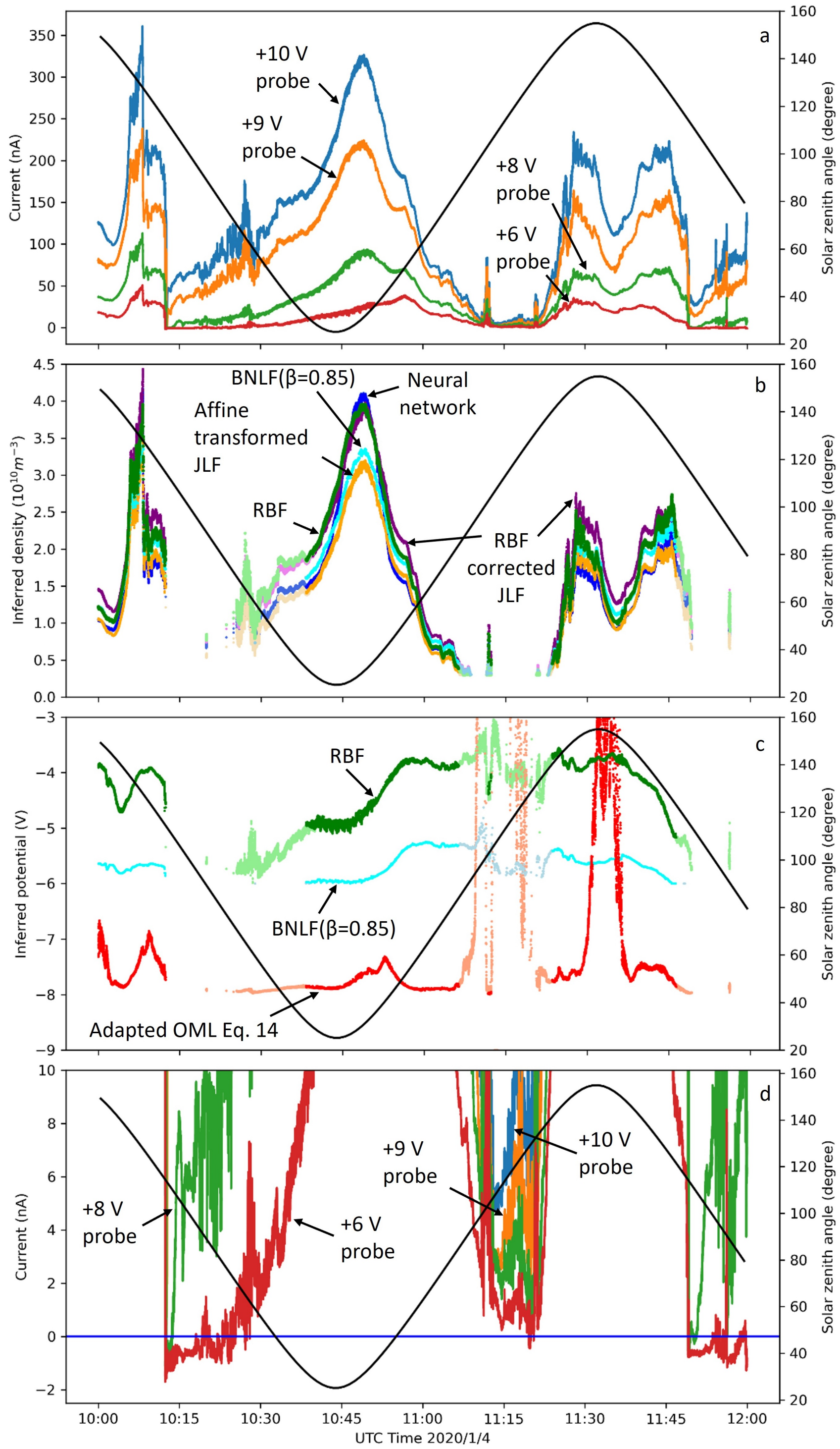

Figure 9. Illustrations of NorSat-1 collected currents considered in this study in panel a, inferred densities in panel b, inferred potentials in panel c, and the NorSat-1 current near 0 A in panel d. The solar zenith angle is also plotted against the secondary axis. Curves in darker colors are from model inferences using data above 10 nA, whereas those in lighter colors show inferences using data with currents between 1 nA and 10 nA.

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. 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.

The densities shown in Fig. 9 panel b are obtained using the five density inference methods mentioned in Sec. 3.1. At 10:45, the neural network density, the RBF corrected JLF density, and the RBF density overlap nicely, while the affine transformed JLF density and the BNLF density ($\beta = 0.85$) are 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 currents and Eq. 14, the inferred satellite floating potential is about -8 V for most of the data range considered in this study as shown in Fig. 9 panel c. This is in stark contradiction with observations in Fig. 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. 14 to infer the satellite potential here, results from the fact that Eq. 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. 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 the generalized OML expression in Eq.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. This potential is generally lower than the potential established by the spacecraft on its own, likely due to the large number of electrons collected by the positively biased solar panels (Ivarsen et al., 2019). 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, and they are consistent with the fact that 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 potential, and to a lesser extent, by the electron temperature. In Fig. 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) peak 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 adapted 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 (Ivarsen et al., 2019). 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. (Ivarsen et al., 2019).

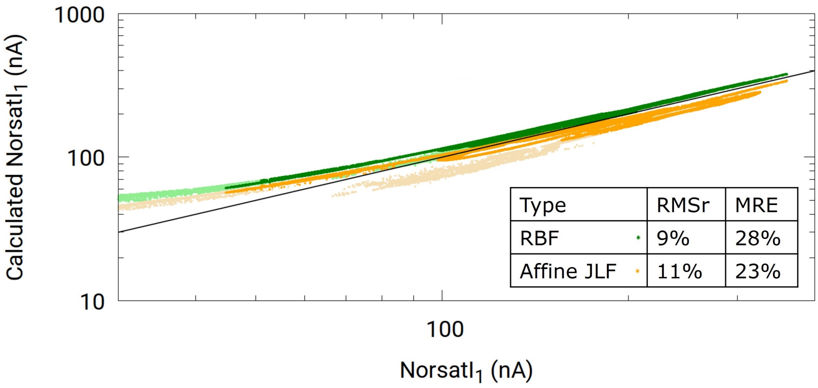

Figure 10. Consistency check is performed in the in situ data following the same procedure as in the synthetic data set. Both models 1 and 2 are trained with our synthetic data, and applied to currents from the +10 V probe on NorSat-1. Darker colors refer to inferences made with currents above 10 nA,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

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. 3.2 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. 10 for the +10 V probe. In 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 using 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 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, confirming that either method should be a significant improvement over the commonly used OML approach based on $\beta = 0.5$. From 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, an adapted OML approach and Barjatya nonlinear fit method. The adapted 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. 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 under 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. In cases where probe characteristics are well approximated with analytic expressions such as Eq. 1, the BNLF technique should prove fast and convenient, as it does not require extensive simulations. Custom simulation-regression models, on the other hand, would require more computational resources, which would necessitate optimization in order to be implemented onboard a satellite. Despite their complexity, however, such models would have the advantage of being more general than models based on fits made with empirical analytic expressions. 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, and the possible scientific payoff.

Acknowledgments

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.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: A system for large-scale machine learning. In (p. 265-283).

Allen, J. E., Boyd, R. L. F., & Reynolds, P. (1957, 3). The collection of positive ions by a probe immersed in a plasma. *Proceedings of the Physical Society. Section B*, *70*, 297-304. Retrieved from `https://doi.org/10.1088/0370-1301/70/3/303https://iopscience.iop.org/article/10.1088/0370-1301/70/3/303` doi: 10.1088/0370-1301/70/3/303

Barjatya, A., & Merritt, W. (2018). Error analysis of multi-needle langmuir probe measurement technique. *Review of Scientific Instruments*, *89*, 1-5. Retrieved from `http://dx.doi.org/10.1063/1.5022820` doi: 10.1063/1.5022820

Barjatya, A., St-Maurice, J. P., & Swenson, C. M. (2013). 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*, *118*, 7316-7328. doi: 10.1002/2013JA018788

Barjatya, A., Swenson, C. M., Thompson, D. C., & Wright, K. H. (2009). Invited article: Data analysis of the floating potential measurement unit aboard the international space station. *Review of Scientific Instruments*, *80*. doi: 10.1063/1.3116085

Bekkeng, T. A., Jacobsen, K. S., Bekkeng, J. K., Pedersen, A., Lindem, T., Lebreton, J. P., & Moen, J. I. (2010, 8). Design of a multi-needle langmuir probe system. *Measurement Science and Technology*, *21*, 085903. Retrieved from https://iopscience.iop.org/article/10.1088/0957-0233/21/8/085903 doi: 10.1088/0957-0233/21/8/085903

Bernstein, I. B., & Rabinowitz, I. N. (1959). Theory of electrostatic probes in a low-density plasma. *Physics of Fluids*, *2*, 112. Retrieved from https://aip.scitation.org/doi/10.1063/1.1705900 doi: 10.1063/1.1705900

Bhattarai, S., & Mishra, L. N. (2017, 8). Theoretical study of spherical langmuir probe in maxwellian plasma. *International Journal of Physics*, *5*, 73-81. Retrieved from http://pubs.sciepub.com/ijp/5/3/2/index.html doi: 10.12691/ijp-5-3-2

Bilitza, D., Altadill, D., Zhang, Y., Mertens, C., Truhlik, V., Richards, P., ... Reinisch, B. (2014, 2). The international reference ionosphere 2012 – a model of international collaboration. *Journal of Space Weather and Space Climate*, *4*, A07. Retrieved from http://www.swsc-journal.org/10.1051/swsc/2014004 doi: 10.1051/swsc/2014004

Chalaturnyk, J., & Marchand, R. (2019, 5). Regression-based interpretation of langmuir probe measurements. *Frontiers in Physics*, *7*. Retrieved from https://www.frontiersin.org/article/10.3389/fphy.2019.00063/full doi: 10.3389/fphy.2019.00063

Chen, F. F. (1965, 1). Numerical computations for ion probe characteristics in a collisionless plasma. *Journal of Nuclear Energy. Part C, Plasma Physics, Accelerators, Thermonuclear Research*, *7*, 47-67. Retrieved from https://doi.org/10.1088/0368-3281/7/1/306https://iopscience.iop.org/article/10.1088/0368-3281/7/1/306 doi: 10.1088/0368-3281/7/1/306

Chen, F. F. (2003). Lecture notes on langmuir probe diagnostics. Mini-Course on Plasma Diagnostics, IEEEICOPS meeting. Retrieved from http://www.seas.ucla.edu/~ffchen/Publs/Chen210R.pdf

Debchoudhury, S., Barjatya, A., Minow, J. I., Coffey, V. N., & Chandler, M. O. (2021, 10). 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*, *126*. doi: 10.1029/2021JA029393

Deca, J., Lapenta, G., Marchand, R., & Markidis, S. (2013). Spacecraft charging analysis with the implicit particle-in-cell code iPic3D. *Physics of Plasmas*, *20*(10), 102902. Retrieved from https://doi.org/10.1063/1.4826951 doi: 10.1063/1.4826951

Ergun, R. E., Andersson, L. A., Fowler, C. M., & Thaller, S. A. (2021, 3). Kinetic modeling of langmuir probes in space and application to the MAVEN langmuir probe and waves instrument. *Journal of Geophysical Research: Space Physics*, *126*. doi: 10.1029/2020JA028956

Frey, P., & George, P. L. (2007). *Mesh generation: Application to finite elements.* ISTE.

Geuzaine, C., & Remacle, J.-F. (2009, 9). Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, *79*, 1309-1331. Retrieved from http://doi.wiley.com/10.1002/nme.2579 doi: 10.1002/nme.2579

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. Retrieved from https://books.google.ca/books?id=Np9SDQAAQBAJ

Guthrie, J., Marchand, R., & Marholm, S. (2021). Inference of plasma parameters from fixed-bias multi-needle langmuir probes (m-NLP). *Measurement Science and Technology*, *32*. doi: 10.1088/1361-6501/abf804

Hoang, H., Clausen, L. B., Røed, K., Bekkeng, T. A., Trondsen, E., Lybekk, B., ... Moen, J. I. (2018, 6). The multi-needle langmuir probe system on board NorSat-1. *Space Science Reviews*, *214*, 75. Retrieved from http://

dx.doi.org/10.1007/s11214-018-0509-2http://link.springer.com/10.1007/s11214-018-0509-2 doi: 10.1007/s11214-018-0509-2

Hoang, H., Røed, K., Bekkeng, T. A., Moen, J. I., Clausen, L. B., Trondsen, E., ... Sagi, E. (2019). The multi-needle langmuir probe instrument for QB50 mission: Case studies of Ex-Alta 1 and hoopoe satellites. *Space Science Reviews*, *215*. Retrieved from http://dx.doi.org/10.1007/s11214-019-0586-x doi: 10.1007/s11214-019-0586-x

Hoang, H., Røed, K., Bekkeng, T. A., Moen, J. I., Spicher, A., Clausen, L. B. N., ... Pedersen, A. (2018, 6). A study of data analysis techniques for the multi-needle langmuir probe. *Measurement Science and Technology*, *29*, 065906. Retrieved from https://iopscience.iop.org/article/10.1088/1361-6501/aab948 doi: 10.1088/1361-6501/aab948

Hoskinson, A. R., & Hershkowitz, N. (2006). Effect of finite length on the current-voltage characteristic of a cylindrical langmuir probe in a multidipole plasma chamber. *Plasma Sources Science and Technology*, *15*, 85-90. doi: 10.1088/0963-0252/15/1/013

Ivarsen, M. F., Hoang, H., Yang, L., Clausen, L. B., Spicher, A., Jin, Y., ... Lybekk, B. (2019). Multineedle langmuir probe operation and acute probe current susceptibility to spacecraft potential. *IEEE Transactions on Plasma Science*, *47*, 3816-3823. doi: 10.1109/TPS.2019.2906377

Jacobsen, K. S., Pedersen, A., Moen, J. I., & Bekkeng, T. A. (2010). A new langmuir probe concept for rapid sampling of space plasma electron density. *Measurement Science and Technology*, *21*. doi: 10.1088/0957-0233/21/8/085902

Johnson, J. D., & Holmes, A. J. T. (1990, 10). Edge effect correction for small planar langmuir probes. *Review of Scientific Instruments*, *61*, 2628-2631. doi: 10.1063/1.1141849

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Y. Bengio & Y. LeCun (Eds.), . Retrieved from http://arxiv.org/abs/1412.6980

Laframboise, J. G. (1966). Theory of spherical and cylindrical langmuir probes in a collisionless, maxwellian plasma at rest. *Toronto Univ. (Ontario). Inst. for Aerospace Studies*. Retrieved from https://www.osti.gov/biblio/4522330

Lebreton, J.-P., Stverak, S., Travnicek, P., Maksimovic, M., Klinge, D., Merikallio, S., ... Salaquarda, M. (2006, 4). The ISL langmuir probe experiment processing onboard DEMETER: Scientific objectives, description and first results. *Planetary and Space Science*, *54*, 472-486. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0032063305002084 doi: 10.1016/j.pss.2005.10.017

Lira, P. A. R., Marchand, R., Burchill, J., & Forster, M. (2019, 8). Determination of Swarm front plate's effective cross section from kinetic simulations. *IEEE Transactions on Plasma Science*, *47*, 3667-3672. Retrieved from https://ieeexplore.ieee.org/document/8725927/ doi: 10.1109/TPS.2019.2915216

Liu, G., & Marchand, R. (2021). Kinetic simulation of segmented plasma flow meter response in the ionospheric plasma. *Journal of Geophysical Research: Space Physics*, *126*. doi: 10.1029/2021JA029120

Liu, G., & Marchand, R. (2022). Inference of m-NLP data using radial basis function regression with center-evolving algorithm. *Computer Physics Communications*, *280*, 108497. Retrieved from https://www.sciencedirect.com/science/article/pii/S0010465522002168 doi: https://doi.org/10.1016/j.cpc.2022.108497

Marchand, R. (2012, 2). PTetra, a tool to simulate low orbit satellite–plasma interaction. *IEEE Transactions on Plasma Science*, *40*, 217-229. Retrieved from http://ieeexplore.ieee.org/document/6069605/ doi: 10.1109/TPS.2011.2172638

Marchand, R., & Lira, P. A. R. (2017, 4). Kinetic simulation of spacecraft–environment interaction. *IEEE Transactions on Plasma Science*, *45*, 535-554. Retrieved from http://ieeexplore.ieee.org/document/7888602/ doi: 10.1109/TPS.2017.2682229

Marchand, R., Miyake, Y., Usui, H., Deca, J., Lapenta, G., Matéo-Vélez, J. C., ... Markidis, S. (2014). Cross-comparison of spacecraft-environment interaction model predictions applied to solar probe plus near perihelion. *Physics of Plasmas*, *21*(6), 062901. Retrieved from https://doi.org/10.1063/1.4882439 doi: 10.1063/1.4882439

Marholm, S., & Marchand, R. (2020, 4). Finite-length effects on cylindrical langmuir probes. *Physical Review Research*, *2*, 023016. Retrieved from https://link.aps.org/doi/10.1103/PhysRevResearch.2.023016 doi: 10.1103/physrevresearch.2.023016

Marholm, S., Marchand, R., Darian, D., Miloch, W. J., & Mortensen, M. (2019, 8). Impact of miniaturized fixed-bias multineedle langmuir probes on CubeSats. *IEEE Transactions on Plasma Science*, *47*, 3658-3666. Retrieved from https://ieeexplore.ieee.org/document/8734022/ doi: 10.1109/TPS.2019.2915810

Mott-Smith, H. M., & Langmuir, I. (1926). The theory of collectors in gaseous discharges. *Physical Review*, *28*, 727-763. doi: 10.1103/PhysRev.28.727

Olowookere, A., & Marchand, R. (2021). Fixed bias probe measurement of a satellite floating potential. *IEEE Transactions on Plasma Science*, *49*, 862-870. doi: 10.1109/TPS.2020.3045366

Rudakov, D. L., Boedo, J. A., Moyer, R. A., Lehmer, R. D., Gunner, G., & Watkins, J. G. (2001). Fast electron temperature diagnostic based on langmuir probe current harmonic detection on DIII-D. *Review of Scientific Instruments*, *72*, 453-456. doi: 10.1063/1.1310577

Saad, Y. (2003). *Iterative methods for sparse linear systems* (Second ed.). Society for Industrial and Applied Mathematics. Retrieved from https://epubs.siam.org/doi/abs/10.1137/1.9780898718003http://epubs.siam.org/doi/book/10.1137/1.9780898718003 doi: 10.1137/1.9780898718003

Sheridan, T. E. (2010, 3). The plasma sheath around large discs and ion collection by planar langmuir probes. *Journal of Physics D: Applied Physics*, *43*, 105204. doi: 10.1088/0022-3727/43/10/105204

Sudit, I. D., & Woods, R. C. (1994). A study of the accuracy of various langmuir probe theories. *Journal of Applied Physics*, *76*, 4488-4498. doi: 10.1063/1.357280