Inference of plasma parameters from fixed-bias multi-needle Langmuir probes (m-NLP)

Inference of plasma parameters from fixed-bias multi-needle Langmuir probes (m-NLP)

Joshua Guthrie[1,*] ⓘ**, Richard Marchand**[1] **and Sigvald Marholm**[2,3]

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

E-mail: jdguthri@ualberta.ca

CrossMark

Abstract
New approaches are presented to infer plasma densities and satellite floating potentials from currents collected with fixed-bias multi-needle Langmuir probes (m-NLP). Using synthetic data obtained from kinetic simulations, comparisons are made with inference techniques developed in previous studies and, in each case, model skills are assessed by comparing their predictions with known values in the synthetic data set. The new approaches presented rely on a combination of an approximate analytic scaling law for the current collected as a function of bias voltage, and multivariate regression. Radial basis function regression (RBF) is also applied to Jacobsen *et al*'s procedure (2010 *Meas. Sci. Technol.* **21** 085902) to infer plasma density, and shown to improve its accuracy. The direct use of RBF to infer plasma density is found to provide the best accuracy, while a combination of analytic scaling laws with RBF is found to give the best predictions of a satellite floating potential. In addition, a proof-of-concept experimental study has been conducted using m-NLP data, collected from the Visions-2 sounding rocket mission, to infer electron densities through a direct application of RBF. It is shown that RBF is not only a viable option to infer electron densities, but has the potential to provide results that are more accurate than current methods, providing a path towards the further use of regression-based techniques to infer space plasma parameters.

Keywords: ionospheric plasma, regression, multi-needle Langmuir probes (m-NLP), 3D kinetic simulations, space

(Some figures may appear in colour only in the online journal)

1. Introduction

Our reliance on space technology requires good first principle understanding of the complex dynamics occurring in our near space environment. Space weather events can affect communications, remote sensing, and scientific satellites in orbit, as well as large power grids and pipelines at Earth surface. Monitoring the state of our space environment is the most basic requirement for understanding and developing reliable interpretive and predictive models. Among the many parameters characterizing the state of a plasma, the density, temperature, and plasma flow velocity are the most fundamental, as they are always affected by changes in our environment, and consequently, they can serve as proxies to infer the state of the ionosphere and magnetosphere. Many instruments have been developed over the years, to measure these parameters. Ionospheric plasmas can be monitored

remotely with ground-based instruments such as incoherent scatter radars (ISR) and ionosondes [26, 27] but such measurements lack the spatial and temporal resolution of *in situ* measurements made with instruments mounted on rockets or satellites. Ground based instruments are generally not mobile, which limits their use to specific regions of space. In comparison, *in situ* measurements made with satellites provide a broader coverage, and higher spatial and temporal resolutions. Among those, Langmuir probes have been the instruments of choice in labs and in space, because of their relative simplicity and the many theories developed to describe their interaction with plasma. Langmuir probes are typically operated in sweep mode, where a bias voltage with respect to a ground is varied periodically between negative, and positive values, resulting respectively in ions and electrons being collected. Parameters such as density or temperature can then be inferred from probe characteristics (current as a function of applied voltage), based on convenient analytic inference procedures. In this study we consider the use of fixed-bias multi-Needle Langmuir probes (m-NLP), first proposed by Jacobsen *et al* to infer plasma density. Many articles have been written on the inference of plasma parameters from Langmuir probe characteristics. Over the years, several different probe geometries have been considered theoretically [1, 24, 25, 41] and experimentally [4, 13, 19], for plasma in different regimes. The approach considered here is to use multi-needle Langmuir probes (m-NLP) with fixed-bias voltages with respect to a spacecraft, to infer the density, and possibly other physical parameters. Compared to sweep-voltage probes, the advantage of fixed-bias Langmuir probes is that they can provide a much higher temporal and, owing to the high speed of spacecraft, spatial resolution. Assuming a sufficiently long and thin probe, the orbital-motion-limited (OML) approximation for probe current collection, implies that the square of the current collected by such a probe should vary linearly with the square of the density times the probe voltage. This led Jacobsen *et al* to propose fixed positively biased multi-needle Langmuir probes as a means of measuring plasma density independently of the temperature or the satellite floating potential. This approach was justified by the fact that electron thermal speeds are much larger than combined satellite speeds in low Earth orbit (LEO), and ionospheric winds. As a result, electrons appear as stationary in a satellite rest frame, and the current collected by a long, positive cylindrical probe is approximated as

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

with $\beta = 0.5$, and where e, n_e, T_e and m_e are the elementary charge, the plasma density, the electron temperature and mass, A is the surface area of the probe, k is the Boltzmann constant, V_b is the probe bias voltage, V_f is the spacecraft floating potential, and $V_f + V_b$ is the probe potential with respect to the background plasma. In equation (1), and in what follows, I is the absolute values of electron collected current. From this expression, it readily follows that I^2 varies linearly with

the bias voltage, with a slope that is independent of the temperature. As a result, the slope K of I^2 as a function of V_b, obtained from two or more currents from probes biased to different voltages, can be used to infer the density with

$$ n_e = \sqrt{\frac{\pi^2 m_e K}{2e^3 A^2}}. \qquad (2) $$

This result has motivated laboratory and rocket experiments, and it led to the use of m-NLPs on several of CubeSats deployed in the QB50 program [12, 37], and to the use of such an array of probes on the larger NorSat-1 satellite [14]. One point to keep in mind when considering equations (1) and (2) however, is that OML equations are derived under somewhat stringent simplifications. In particular it is assumed that (a) the probe radius is much smaller than the plasma Debye length, (b) plasma is unmagnetized, (c) the background plasma velocity distribution function is Maxwellian, (d) plasma flow speed is negligible compared to the thermal speed of the species being collected, and (e) the probe is much longer than the Debye length in order for end effects to be negligible. For probe diameters of order 0.5 mm or less, assumption (a) is satisfied. The thermal gyroradius of electrons in the ionosphere being of order 1 cm or more, (b) is also satisfied. Frequent collisions with neutrals cause electron velocity distributions to be nearly Maxwellian at low and mid latitudes, where (c) is satisfied. As for electron thermal speeds, they are typically more than an order of magnitude larger than low Earth orbit satellite speeds, so that (d) is well satisfied as well. The problem with equation (2) stems from the use of the OML theory, in which probe lengths are assumed to be sufficiently long for end effects to be negligible. Indeed as demonstrated experimentally [19, 41], and theoretically [22, 23, 33], probe lengths have to be much larger than the Debye length, by factors ranging between several tens, to hundreds, for the OML approximation to be applicable with $\beta = 0.5$. This condition, however, is generally not satisfied with needle probes mounted on CubeSats or NorSat-1 in ionospheric plasma. As a result, while equation (1) can still provide a good scaling law for the collected current as a function of voltage, the value of β is no longer 0.5, and equation (2) no longer provides an accurate estimate of the density. For finite-length probes, in which end effects contribute to the collected current, β is found experimentally and theoretically to range between 0.5 and 1.0, and the technique used to infer the density must be modified accordingly.

In response to this predicament, two solutions were proposed by Hoang *et al* [16], and Barjatya *et al* [2], in order to improve the inference of the plasma density, while accounting for the fact that β may be different from 0.5. In both studies, the scaling law given in equation (1) is assumed, but β is now treated as a parameter to be determined. In their approach, Barjatya *et al* use a nonlinear fit to determine the unknown parameters n_e, V_f, T_e, and β appearing in equation (1), in order to match the currents collected by the four probes in the m-NLP. Similar fits had previously been used by Barjatya *et al*

to determine β, the floating potential, and the electron density, from the Floating Potential Measurement Unit attached to the International Space Station [3]. While the determination of four parameters from four independent measurements is possible in principle, it was found that the determination of the temperature could not be made accurately in this straightforward approach. Noting that the inference of other parameters was relatively insensitive to T_e, the solution proposed consisted of specifying an approximate value for the temperature, and then using a nonlinear fit to determine the remaining three parameters from the currents collected by the probes with the three largest bias voltages. This was justified by the fact that, based on synthetic data generated with equation (1), assuming a range of β values between 0.5 and 0.65, significantly more accurate inferences of the density were made than with Jacobsen's original technique (that is, assuming $\beta = 0.5$) even if the temperature used in the nonlinear fit was varied by $\pm 100\%$ relative the actual temperature used to generate the data set.

Alternatively, Hoang *et al* assessed two approaches; the first one being based on Jacobsen's original least squares linear fit to multiple collected currents assuming $\beta = 0.5$, and a second one, based on different implementations of nonlinear fits, similar to those considered in [2]. With nonlinear fits, four or three currents with different relative weights are considered to determine β, V_f, and n_e. As in [2], it was noted that nonlinear fits were relatively insensitive to the temperatures assumed in equation (1). Temperature values are nonetheless needed in order to perform nonlinear fits and obtain good accuracy for the inferred parameters. In space, it was suggested that temperature estimates be obtained from the International Reference Ionosphere (IRI) model [5] or incoherent scatter radar measurements. The article also reports comparisons of inferred densities with those obtained from the IRI model, and independent measurements in a laboratory plasma [10]; both being deemed satisfactory, and constituting an improvement over estimates made with Jacobsen's original technique.

In the following we present yet two alternative approaches based on a combination of analytic approximation, and multivariate regression, for which inference skills are assessed, using a synthetic data set obtained from simulations, as well as from actual data from a rocket mission. While model validation should ideally be made with actual measurements, the advantage of synthetic data is that it enables the assessment of predictive models with known plasma densities, temperatures, and floating potentials. Constructing such data sets also avoids biasing the data set which would result, for example, from an assumed analytic expression for the scaling of current as a function of voltage. In the next section we explain how kinetic simulations were used to construct a synthetic data set, and present our two inference approaches. In section 3, inference skills are assessed for each of these two approaches, and compared with those of previous models. The application of radial basis function (RBF) regression to experimental data collected from the Visions-2 [39] sounding rocket is presented in section 4. Finally, a summary of our findings and concluding remarks are presented in section 5.

2. Methodology

In order to train and validate inference models as those from m-NLP, it is necessary to have data sets with low level (L1B in satellite data parlance) currents and associated plasma and satellite parameters (n_e, T_e, V_f). Ideally such a data set should be constructed from actual accurate measurements, cross-validated with different instruments, but this is rarely possible in practice, owing to challenges in making such measurement in space or lab plasma [13, 40–42]. An alternative is to use synthetic data sets using analytic models [2, 7], or computer simulations [31, 33], from which precise values of collected currents and corresponding plasma parameters are known. Admittedly, computer simulations, while more accurate than analytic models, do not account for all processes at play in an experiment or in space. They nonetheless make it possible to construct self-consistent data bases from which inference models can be tested and predictive skills quantified. Data sets constructed from simulations are also free from bias which would result from using analytic expressions. This is particularly important, when these same expressions and scaling laws are used to construct and asses inference predictive models. Given data sets, the next step is then to construct and validate models capable of inferring plasma parameters from measured currents. These procedures are described in detail below.

2.1. Data sets

Two distinct synthetic data sets are constructed and used in our model skill assessments. The first one uses the three-dimensional particle-in-cell (PIC) simulation code PTetra to simulate a needle probe in a flowing plasma, with velocity perpendicular to the probe axis. In PTetra, the simulation domain consists of an unstructured adaptive tetrahedral mesh in which Poisson's equation is solved at each time step using Saad's GMRES sparse matrix solver [36]. The validity of simulation results obtained with PTetra has been assessed in previous publications, in which results were compared with theory [28], and those obtained with independently developed computer models [8, 29, 30]. More information about the code can be found in [28, 29]. The probes simulated have the same radius, $r = 0.255$ mm, as the ones on the QB50 CubeSats, but they are twice as long with a length $L = 50$ mm instead of 25 mm [17]. Longer probes would have the advantage of collecting more current with higher signal-to-noise ratio, while being less affected by end effects. The probe dimensions and plasma parameters used in the simulations are given in table 1. Variations in the plasma flow speed by 1 km s^{-1} in different directions relative to the probe, as well as different ion compositions have been considered in selected cases, which resulted in only minor effects on collected currents. This is why we limited our study to the parameters listed in table 1. Simulations are made for 25 combinations of densities and temperatures (five densities, each with five temperatures), and we consider four probes biased to 2, 3, 4 and 5 V with respect to the spacecraft. However, since the spacecraft floating potential varies, it is necessary to obtain collected currents for a

Table 1. Probe dimensions and plasma environment conditions assumed in PTetra simulations. All probes are on the ram side of the satellite, and oriented perpendicularly to the ram direction.

Probe length	50 mm
Probe radius	0.255 mm
Densities	$10^{10}, 3 \times 10^{10}, 10^{11}, 3 \times 10^{11}, 10^{12} \mathrm{m}^{-3}$
Temperatures ($T_e = T_i$)	$0.07, 0.095, 0.12, 0.145, 0.17$ eV
Ion species	100% O^+
Plasma flow speed	7672 km s^{-1}
Floating potentials	-2 V to $+2$ V in steps of 0.2 V
m-NLP bias voltages	2, 3, 4, 5 V

large number of different probe voltages from the simulations. To reduce the number of PIC simulations, only a few (at least four) probe voltages V with respect to the background plasma, in the range 0–7 V with respect to the background plasma, are in fact simulated for each combination of density and temperature. The resulting currents are then fitted to the analytic expression

$$I = a \left(b + \frac{eV}{kT_e} \right)^c \qquad (3)$$

in order to enable the calculation of collected current for arbitrary bias and floating voltages in the range 0 V $\leqslant V = V_f + V_b \leqslant 7$ V. This analytic expression involves three adjustable parameters a, b, c, which are determined from the four or more currents obtained from simulations, and calculated with a nonlinear least squares fit, with differential evolution [9] as the optimization algorithm. Clearly, equation (3) is very similar to equation (1) found in the OML approximation for a probe of infinite length. The parameter b however has no counterpart in equation (1), and it was introduced to relax possible bias in the fits with the expression resulting from the OML approximation, and assumed in some of the inference approaches considered below. This parameter is introduced to account for the fact that equation (1) is not exact, even if it is generally a good approximation for the collected current. By setting the fitting parameter b to unity, we would constrain the interpolation of our simulation results to have exactly the form prescribed in equation (1), while our goal here is to have the best analytic fit for currents computed in our simulations. In practice, depending on the parameters, our fits produce values for b ranging from approximately 0.7 to 1.1. We remark that our simulations do not account for a guard cylinder that would be supporting the probe, and be set to the same potential. For that reason, in order to approximate the effect of a guard, the probes are subdivided into five segments of equal lengths (10 mm each), and the current collected by one of the end segments is replaced by the current collected by the segment next to it. A correlation plot of fitted currents as a function of actual (simulation) collected currents in figure 1 shows the excellent agreement between fits and data. In most cases, the relative error in the fits is of order 1%, and the maximum relative error among all the cases, is under 4%. Given coefficients a, b and c for each of the 25 combinations of densities and temperatures, it
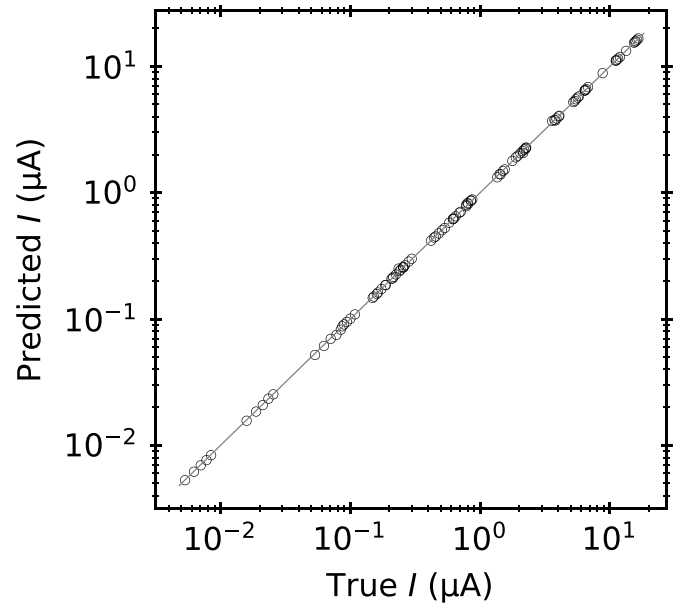
Figure 1. Correlation plot of the fitted current as a function of the actual simulation current. The solid line represents perfect correlation.

is then possible to construct a data set with 4-tuples of currents corresponding to 4-tuples of bias voltages V_b and arbitrary floating potentials V_f in a range such that $0 \leqslant V_f + V_b < 7$ V. Several increments have been tried between successive bias voltages, ranging between 0.75 to 1.5 V, and found to have relatively little impact on prediction accuracy. In this data set, bias voltages of 2, 3, 4, and 5 V are considered, with 21 uniformly distributed floating potentials in the range $(-2, 2)$ V; thus forming a set of $25 \times 21 = 525$ entries or nodes. Thus, each entry in the data set consists of a 4-tuple of currents, along with associated density, temperature and floating potential.

The second data set considered is constructed using the Langmuir software [32], which uses the fits reported in [33]. In this article, fits were constructed for a thin cylindrical probe in a wide range of non-dimensionalised plasma parameters. These can be used to predict the current per unit length along a probe, as well as the total current collected, for different ratios of probe length to the Debye length, and ratios between the probe voltage to the electron temperature. Prescriptions were also derived to approximate the effect of a guard, which would reduce or eliminate end effects on one end of a probe. The simulation results used for the fits in [33] were obtained using PTetra. Since the work in [33] made similar assumptions to those in the OML theory, except for the finite length of the probe, one would expect the results of Langmuir to approach those of OML for a cylinder as the probe length is increased. This is indeed the case, the worst-case discrepancy being less than 5% [32]. Further on, as the probe is shortened to less than the Debye length, one expects the collected current to be proportional to the probe voltage, similar to a spherical probe (though the exact current may not be known, since the effective spherical surface area may differ from the true surface area). This is also observed in [33], where the current–voltage

characteristics for a probe shorter than the Debye length fits a function similar to (1), but with $\beta \sim 1$. This model is used to generate a data set consisting of 10 000 4-tuples of currents for randomly distributed temperatures, densities, and floating potentials in the same range in parameter space, as assumed in the first data set. The temperature and floating potential are uniformly distributed within their ranges, whereas the density is logarithmically distributed (i.e. $\log n_{\mathrm{e}}$ is uniformly distributed). The bias voltages are also the same as in the first data set. Contrary to the first data set, however, this one does not account for a plasma flow, which is deemed negligible, due to the small drift velocity compared to electron thermal speed. Moreover, the data set is generated assuming an ideal (infinite) guard on one end of the probes.

2.2. Inference models

We now turn to the construction of models capable of inferring plasma parameters from measured currents. In addition to the methods already mentioned in the introduction, with which comparisons will be made, two approaches are presented. In model 1, parameters are obtained in part from the assumed relation between current and voltage in equation (1), and in part from multivariate regression. In model 2, inference is made directly using multivariate regression, without relying on any analytic scaling law between current and voltage. Since both models 1 and 2 make use of RBF regressions, we start with a brief presentation of the method.

2.3. Radial basis function regression (RBF)

Multivariate regression offers a general means of inferring dependent variables from scattered data in a multidimensional space. Among the several possible approaches, RBF was chosen for its relative simplicity and accuracy [6, 20, 35]. The method consists of a weighted superposition of functions of the 'radial' distances between points in a multidimensional space, where regression is to be made. Given a set of N reference nodes, or 'pivots' $\{(X_i, Y_i), i = 1, N\}$, where X_i and Y_i are respectively independent and dependent vectors, and assuming an L^2 norm, or Euclidean metric for the distance between two points in X space, RBF regression consists of approximating Y for an arbitrary X, as

$$Y \simeq \sum_{j=1}^{N} a_j G(|X - X_j|), \tag{4}$$

where a_j are regression coefficients, and G is a radial basis function. Regression coefficients a_j can then be determined by requiring exact collocation at pivots; that is, by solving the set of linear equations

$$Y_i = \sum_{j=1}^{N} a_j G(|X_i - X_j|). \tag{5}$$

The choice of the G function is arbitrary, with the only constraint that the equations in equation (5) be linearly independent from one another. The construction of an RBF regression

model is done in two steps. The interpolation function and pivots are first set so as to best approximate dependent variables in a 'training set' in which X and Y are known. The trained model is then applied to a distinct 'validation set', not used in training, and covering the same range in parameter space. In each case, model prediction skill is assessed with a 'cost function' $\mathcal{C}$, which vanishes if predictions match data values exactly, and increases with increasing discrepancies. Several functions have been tried for training, and $G(x) = x^{1.8}$ is used throughout because of the good results that it produces in our problem. Given a function G, the choice of pivots is critical in order to construct an accurate model. In our analysis, training and validation sets are subsets of a larger set, or solution library, constructed from kinetic simulations, as described in section 2.1. In training, given a function G, the objective is to distribute pivots in order to obtain the highest accuracy when applying the model to a data set in which both X and Y values are known. Different approaches have been proposed to achieve this task, including k-clustering [18], and Gaussian clustering [34]. Assuming a number N of pivots, and a number $\mathcal{N}$ of nodes in the training set, we adopt a straightforward strategy, consisting of trying all possible combinations of N pivots among $\mathcal{N}$ nodes in the training set, and selecting the distribution of pivots for which $\mathcal{C}$ is minimum. Two cost functions are used in this study, depending on the nature of the physical parameter being modeled. For the density, which varies over two orders of magnitude, we use the maximum relative error between predictions and actual values in a given data set. For the floating potential, which can vary continuously between negative and positive values, $\mathcal{C}$ is the maximum absolute error between predicted and actual values.

2.4. Model 1: analytic-regression based

The first model considered consists of three steps in which (a) the exponent β, (b) the floating potential and the temperature, and (c) the density are successively estimated.

The starting point in (a) is the empirical relation between collected current and voltage in equation (1). By raising each side of the equation to the power $\alpha = 1/\beta$, we obtain

$$I^{\alpha} = \left(n_{\mathrm{e}} e A \sqrt{\frac{kT_{\mathrm{e}}}{2\pi m_{\mathrm{e}}}} \right)^{\alpha} \frac{1}{kT_{\mathrm{e}}} \left(kT_{\mathrm{e}} + e(V + V_{\mathrm{f}}) \right). \tag{6}$$

Given two currents I_1, I_2 collected by two probes biased to voltages V_1, and V_2, it is straightforward to solve for $V_{\mathrm{f}} + kT_{\mathrm{e}}/e$ in terms of the two voltages and currents and obtain

$$V_{\mathrm{f}} + \frac{kT_{\mathrm{e}}}{e} = \frac{V_2 I_1^{\alpha} - V_1 I_2^{\alpha}}{I_2^{\alpha} - I_1^{\alpha}}. \tag{7}$$

Now, if a third probe is used, with bias voltage V_3, collecting current I_3, the following identity must be satisfied:

$$\frac{V_2 I_1^{\alpha} - V_1 I_2^{\alpha}}{I_2^{\alpha} - I_1^{\alpha}} - \frac{V_3 I_2^{\alpha} - V_2 I_3^{\alpha}}{I_3^{\alpha} - I_2^{\alpha}} = 0, \tag{8}$$

since T_{e} and V_{f} are constants, independent of the bias voltages or currents. In this equation, only α is unknown, because the

currents are measured, and the bias voltages are set by design of the instrument. It is then straightforward to solve for α, and hence β, using a standard numerical root finder.

In step (b), given β, neglecting kT_e/e, which is generally small compared to V_f, equation (7) is used to make a first estimate of the floating potential, as

$$V_{f1} \simeq \frac{V_3 I_2^\alpha - V_2 I_3^\alpha}{I_3^\alpha - I_2^\alpha} \qquad (9)$$

where subscript 1 is used to label this first approximation of V_f. This first estimate of V_f can now be improved by regression to approximate the error in our first inference, $\delta V_{f1} = V_{f1} - V_f$. This is carried out with RBF, using 4-tuples of measured currents as input, and as output, the known difference $\delta V_{f1} = V_{f1} - V_f$ between the first inference V_{f1}, and the floating potential in our data set. This inferred correction is then used to construct a second inference of V_{f2} with improved accuracy. Referring to equations (7) and (9), it is seen that the model of the correction δV_{f1} also provides an estimate of the electron temperature: $kT_e/e \sim \delta V_{f1}$.

Finally in step (c), given the estimates β, V_{f2}, and T_e, it is possible to make a first inference of the density n_{e1} analytically from equation (1). As a final improvement, the relative difference between n_{e1} and the known value from our data set, $\delta n_{e1} = (n_{e1} - n_e)/n_{e1}$ is modeled with RBF, again with the 4-tuples of currents as input. The modeled correction is then applied to n_{e1} to yield a further improved density estimate n_{e2}. In practice, the increase in accuracy between n_{e1} and n_{e2} is only modest, but n_{e2} is found to be better centered around the exact values. For that reason, only inferred n_{e2} is considered below.

To summarize, model 1 involves several steps consisting of analytic and regression estimates, from which the four parameters β, V_f, T_e, and n_e are estimated from 4-tuples of currents obtained with four given bias voltages as the only input. It is noted that the procedure involving equations (8) and (9) only requires 3-tuples of currents and bias voltages. The RBF corrections to V_f and n_e however, are done using the four currents and voltages, owing to the fact that four parameters (β, V_f, T_e, and n_e) need to be determined. Results obtained with both three and four sets of collected currents are presented in section 3.3 below.

2.5. Model 2: direct RBF regression

In this approach, RBF regression is used to directly infer physical parameters, without relying on any analytic approximation for the currents and bias voltage, as in method 1. The advantage here is that the resulting models are unbiased relative to any approximate theory; that is, they are purely data-driven. The added challenge however, is that they must accurately reproduce the full dependence of parameters such as densities and satellite floating potentials, directly from 4-tuples of collected currents.

## 3. Assessment of model inference skills

In this section, the models proposed independently by Barjatya *et al* [2], and Hoang *et al* [15], and models 1 and 2 described in 2.4 and 2.5 are assessed with data sets obtained from kinetic simulations as described in section 2.1. For models 1 and 2, 200 randomly selected nodes are used for training, and the remaining 325 nodes are used for validation.

The models are assessed using different skill metrics. For the floating potential we define the error of a data point as $V_{f,model} - V_{f,actual}$, and compute the maximum absolute error (MAE) and the root mean square of the error (RMSE). In addition, we decouple $RMSE^2 = \mu^2 + \sigma^2$ into a bias/offset μ and a (population) standard deviation σ of the error to more clearly identify to which extent the error is caused by a systematic offset or a less predictable spread. The offset μ is calculated as the average error, and σ is the standard deviation of the error in a given data set. For the density, because it spans several orders of magnitude, we use the relative error, defined as $(n_{e,model} - n_{e,actual})/n_{e,model}$. This is chosen rather than the more usual definition where relative errors are with respect to exact values, because, from an operational point of view, model inference is made for variables that are not known otherwise. It is therefore more convenient to assess margins of uncertainty with respect to prediction values, which are known, than with exact values, which are not known. With this, we report the maximum (absolute value of) the relative error (MRE) and the root mean square of the relative error (RMSrE) for the density. Again, we decompose RMSrE into the bias/offset μ_r, and standard deviation σ_r of the relative error. For consistency with past literature, we also report the Pearson correlation coefficient R for both the density and floating potential, although R is known to be close to unity even for relatively large errors [2].

3.1. Jacobsen et al's linear fit approach, with $\beta = 0.5$

Using the linear fit approach proposed in [21], and summarized in section 1, densities are inferred from the 4-tuples of currents in our solution library constructed from kinetic simulations described in section 2.1. The correlation plot of these results is shown in figure 2, with selected skill metrics. The linear fit inference is seen to significantly overestimate densities, by factors ranging from 3 to 9, relative to densities in our data set. It is interesting to note that each vertical cluster in the figure consists of $21 \times 5 = 105$ circles, which is the number of combinations in floating potentials and temperatures considered in the construction of the data set. It follows that the spread in inferred density, using this approach, is mainly caused by the spread in floating potentials and temperatures, which are not accounted for in this linear fit formalism.

The regularity in the discrepancies between inferred and data base densities, however, suggests that it should be possible to improve model predictions in this case, using regression. Thus, RBF was used to construct a model for the relative difference between predicted and data densities seen in figure 2. Three pivots were found to be optimal in this case, to shift the centroid of predicted densities close to the ideal
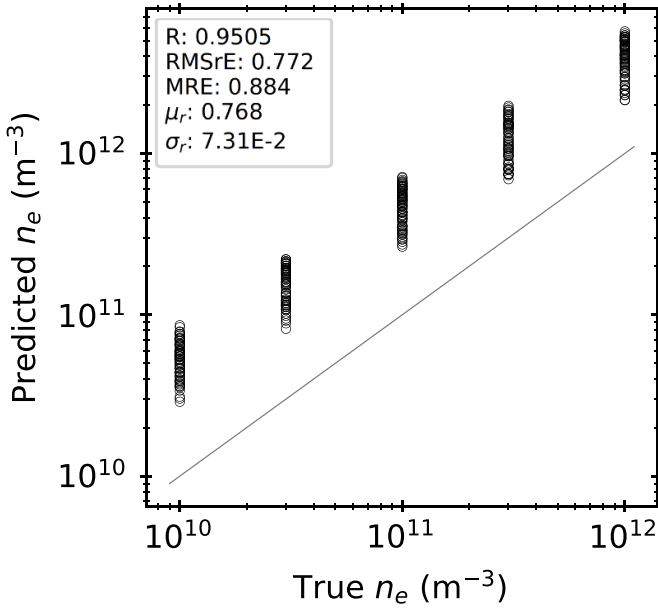
Figure 2. Correlation plot of the density obtained with Jacobsen *et al*'s linear fit.
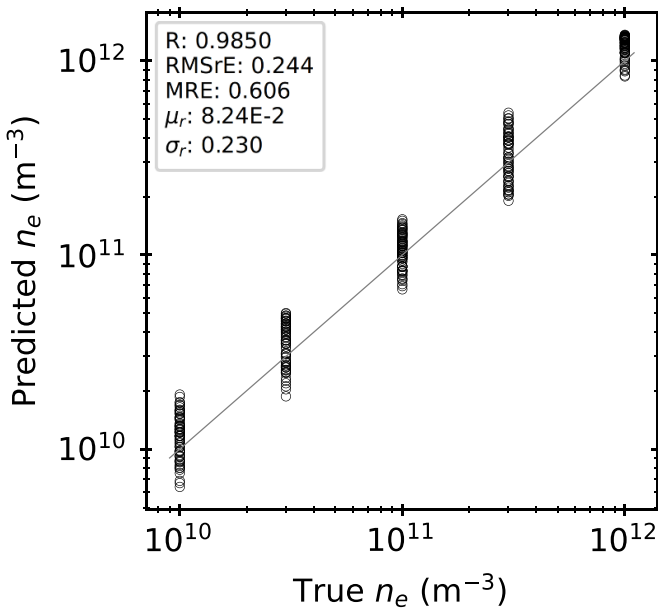
Figure 3. Correlation plot of the density obtained with Jacobsen *et al*'s fit, corrected with RBF regression using three pivots.

correlation curve, without over-fitting. This model was then applied to the validation set to assess the skill. As shown in figure 3, while the spread in the vertical clusters of circles remains significant, the centroid of the predicted densities is now much closer to the ideal correlation curve; resulting in a notable improvement in the skill metrics.

3.2. Nonlinear least squares fits

Nonlinear least squares, similar to those proposed by Barjatya *et al* and Hoang *et al* summarized in section 1 are now considered for determining the floating potential and electron

density. This is done by considering three possible implementations consisting of (a) a 4-parameter fit using all four currents, (b) a 3-parameter fit using currents from the three largest bias voltages (to account for the possibility of a probe with negative voltage), and (c) a 3-parameter fit using currents from all four probes. In all cases, nonlinear least square fits are made using the Python library SciPy, to perform differential evolution optimization [38]. In both cases (b) and (c), the temperature appearing in equation (1) is set to the exact value from the solution library, in order to reduce the number of fitting parameters from four to three. This is similar to the approach taken by Barjatya *et al* who used estimated values of the temperature, and by Hoang *et al* who used estimates from the International Reference Ionosphere (IRI) [5], and EISCAT measurements [11], in their fits. By setting the temperature to its true value, these results produce the best possible fits with these approaches. Model skills are summarized in tables 2 and 3 for each case. In case (a), consistently with findings from [2], fitted temperatures are found to be very inaccurate, which explains the lower performance of the four-parameter fit approach compared to the other two. From the tables, cases (b) and (c) are seen to result in nearly identical skills, although three parameter fits from the four probe currents (case (c)) is found to be slightly more accurate. Inference skills obtained in case (c) are shown in figures 4 and 5 for the floating potential and density, respectively. Excellent agreement is seen for inferred V_f compared to known values from our validation set, with a maximum absolute error of 0.101 V in the range $(-2, 2)$ V of possible floating potentials. Densities are also modeled with good accuracy, with a maximum relative error of approximately $\pm 59\%$ over the $(10^{10}, 10^{12})$m^{-3} range.

3.3. Model 1: analytic-regression based

Following the procedure outlined in section 2.4, which consists of several steps involving a root finder, analytic expressions, and regression, models were constructed with synthetic training and validation sets obtained from simulations. Correlation plots computed with the validation sets are shown in figure 6 for the floating potential V_{f2}, and figure 7 for the density n_{e2}. Referring to tables 2 and 3, inferred floating potentials are seen to be slightly less accurate compared to those obtained from three-parameter fits, while inferred densities are more accurate. We must recall however, that in the three-parameter fits, known temperatures from the data sets were used, while in model 1, the temperature is one of the fitting parameters, and that its determination is rather inaccurate. Thus, in order to have a fair comparison between model 1 and the three-parameter fit approaches, model 1 inferences were made in which the known temperatures were used. The results in the tables show that model 1 predictions of the floating potential are now as accurate as those from the three-parameter fits, but that the density predictions are significantly more accurate. For either V_f or n_e, model 1 predictions, with or without specifying known values of T_e, are also significantly more accurate than those obtained from a four-parameter fit, which would be required should an accurate measurement of the temperature not be available.

Table 2. Quantitative skill test results for the floating potential using each of the data analysis methods described. R refers to the Pearson correlation coefficient, RMSE to the root mean square error, MAE to the maximum absolute error, μ to the offset, and σ to the standard deviation in the error.

Data analysis method	R	RMSE	MAE	μ	σ
4 Parameter nonlinear least squares fit	0.9975	0.0983	0.247	−0.048	0.086
3 Parameter nonlinear fit, 4 probes	0.9995	0.0398	0.101	0.0035	0.040
3 Parameter nonlinear fit, 3 probes	0.9995	0.0398	0.101	0.0035	0.040
Model 1 with known values of T_e	0.9995	0.0398	0.101	0.0034	0.040
Model 1 with inferred values of T_e	0.9995	0.0409	0.130	−0.0091	0.040
Model 2—Direct RBF	0.9970	0.0970	0.361	−0.026	0.093

Table 3. Quantitative skill test results for the electron density using each of the data analysis methods described. R refers to the Pearson correlation coefficient, RMSrE to the root mean square of the relative error, MRE to the maximum relative error, μ_r to the offset, and σ_r to the standard deviation in the relative error.
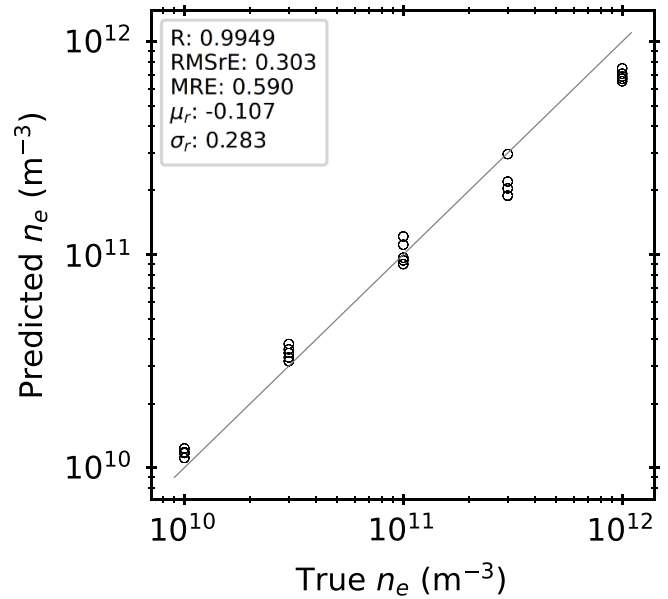
Data analysis method	R	RMSrE	MRE	μ_r	σ_r
Linear fit	0.9505	0.77	0.88	0.77	0.073
Linear fit with RBF correction	0.9850	0.24	0.61	0.082	0.23
Linear fit, `Langmuir` data set	0.9934	0.54	0.75	0.521	0.12
4 Parameter nonlinear fit, 4 probes	0.9819	0.28	1.23	−0.018	0.28
3 Parameter nonlinear fit, 4 probes	0.9949	0.30	0.59	−0.11	0.28
3 Parameter nonlinear fit, 3 probes	0.9949	0.30	0.59	−0.11	0.28
Model 1 with known values of T_e	0.9944	0.14	0.32	−0.0055	0.14
Model 1 with inferred values of T_e	0.9954	0.13	0.27	0.035	0.13
Model 2—Direct RBF	0.9988	0.075	0.18	−0.0092	0.074
Model 2—Direct RBF, `Langmuir` data set	0.9962	0.075	0.21	−0.0038	0.075
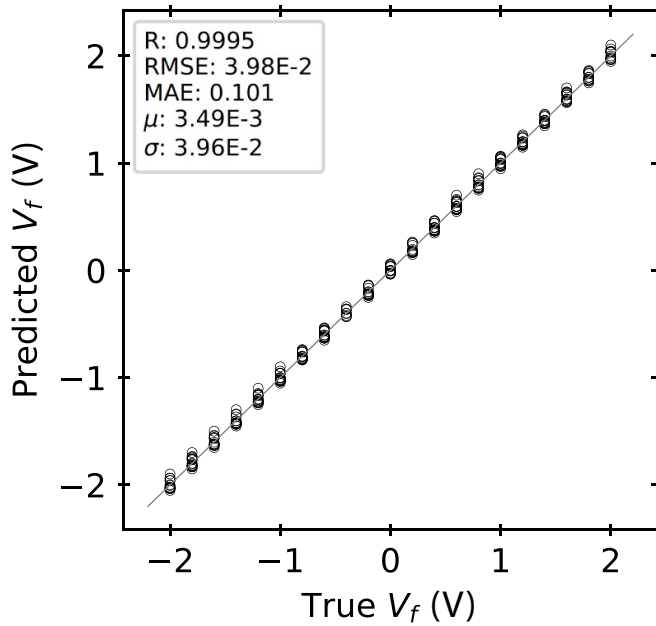
Figure 4. Correlation plot for the floating potential obtained with the 3-parameter nonlinear least squares fitting approach and all four probe currents. In the fits, temperatures are set to the exact value appearing in our solution library.

Figure 5. Correlation plot for the density obtained with the 3-parameter nonlinear least squares fitting approach and all four probe currents. In the fits, temperatures are set to the exact value appearing in our solution library.
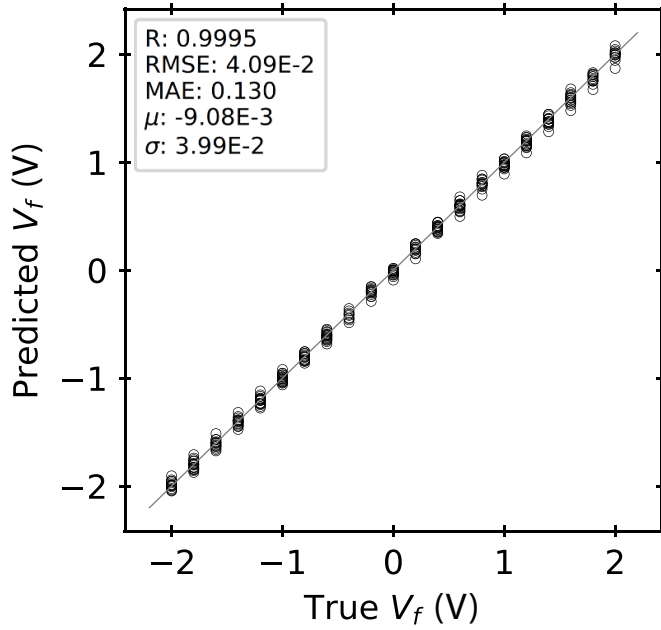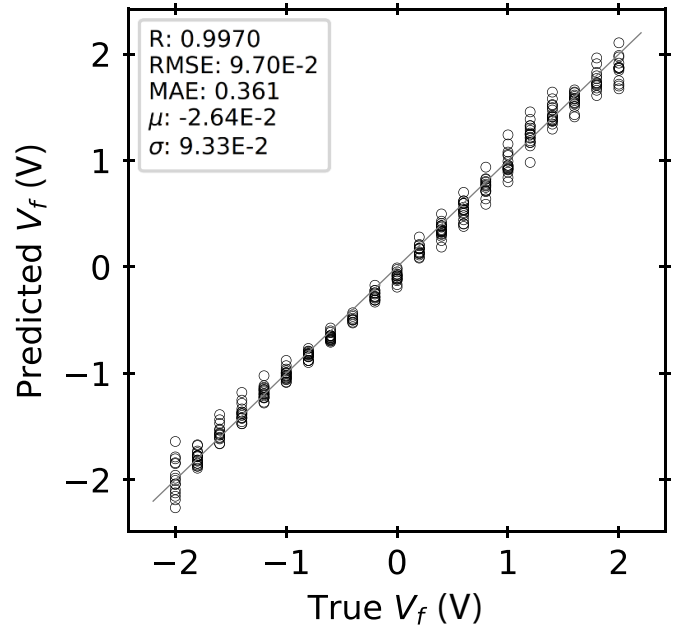
3.4. Model 2: direct RBF regression

The most straightforward model consists of using RBF regression directly to infer the floating potential and plasma density. This is done without any intermediate analytic steps, which results in models that are fully data-driven, and unbiased to analytic approximations. Considering that in this approach regression is expected to reproduce the full dependence of the variables of interest from 4-tuples of currents, as opposed to small corrections to estimates obtained by other means, it is not clear *a priori*, whether inferences should be more or less
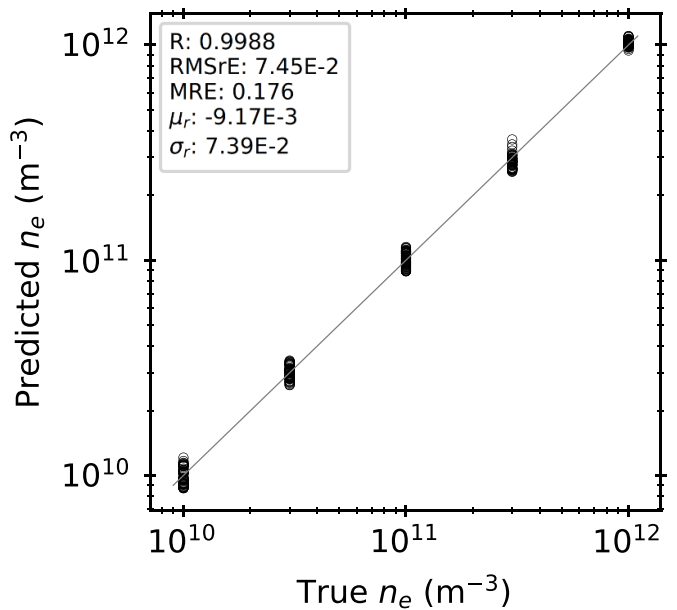
Figure 6. Correlation plot of the floating potential obtained with model 1 and RBF estimates of the temperature using four pivots.

Figure 8. Correlation plot of the floating potential obtained with direct RBF regression, with five pivots. In this case, 4-tuples of currents are normalized by dividing by the largest collected current.
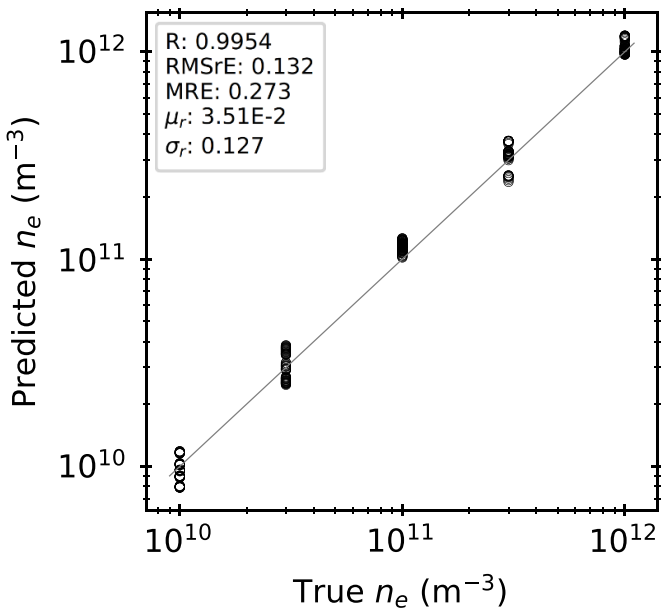
Figure 7. Correlation plot of the density obtained with model 1 and RBF estimates of the temperature using four pivots.

Figure 9. Correlation plot of the density obtained with direct RBF regression, with five pivots.

accurate than those of method 1. Correlation plots are shown in figures 8 and 9 for inferred floating potentials and densities, respectively, using RBF with five pivots. While model predictions of V_f follow the ideal correlation line in figure 8, with nearly the same slope and cluster centroids close to the ideal correlation line, their vertical spread is larger than in plots of V_f from other models, and prediction skills are seen to be the lowest among all models considered. The situation is opposite for the density however, for which inferred densities show the best agreement with those from the validation data set. This is also clear from the skill metrics listed in table 2, which are the best

among the eight models considered. An interesting observation is that, with our training and validation data sets, the direct RBF approach produces the highest accuracy for the density, and the lowest one for the floating potential. This shows that different approaches may be better adapted to model different physical parameters. As a final remark, very little has been said so far about modeling the temperature. The reason is that all attempts have produced very scattered and inaccurate estimates of the temperature, whether with four-parameter fits, model 1, or with direct RBF. This is consistent with findings

reported by Barjatya *et al* and Hoang *et al* and it is a consequence of the relatively weak dependence of collected currents on the temperature for these types of probes.

4. Application to other data sets

In this section method 2 is further tested by inferring densities in two data sets constructed independently from the one considered in the previous sections. The focus here on method 2 and density predictions is motivated in part by the original purpose of using m-NLPs, to infer densities independently of temperatures. Another reason is that method 2 is independent of *a priori* analytic expressions such as equation (1), which makes it applicable under more general conditions, in which equation (1) may not be a good approximation. The first set is constructed with the `Langmuir` program, in which the current collected by a probe is interpolated from numerically computed currents on a grid of non-dimensionalised plasma and probe parameters [32]. The second, experimental, consists of 4-tuples of currents measured in the Visions-2 rocket mission, with densities inferred using two independent techniques.

4.1. Blind test with `Langmuir` generated data

To better assess method 2, it has been applied to a second data set in a blind study, where one author (SM) assessed the skill of predictions made by the other authors. As mentioned in 2.1, this second data set consists of 10 000 4-tuples of currents for different plasma parameters. 200 of these were used for training, and yet another 800 were made available for quick assessments and experimentation during the training phase. The true plasma parameters behind the remaining 9000 4-tuples of currents were not seen by the experimenters, and were only used later to compute skill metrics of the predictions by the last author.

Correlation plots for predicted densities are shown in figure 10, with corresponding skill metrics included in table 3. With `Langmuir`, currents are calculated for a probe geometry, and plasma conditions different from what was assumed in section 2.1. The excellent correlation between given and inferred densities, with similar metrics to those seen in figure 9, is promising and provides strong support to the applicability of the method to experimental data. For comparison, predictions of the density from the `Langmuir` data set, using a Jacobsen *et al*'s linear fit with $\beta = 0.5$ is also included. Compared with the linear fit, which largely overestimates the density with a 52% bias, model 2 predictions have the lowest bias and standard deviation; thus providing a significant improvement to the predicted density.

### 4.2. Application of RBF to visions-2 experimental data

Method 2, our best density inference technique, is also tested against experimentally inferred densities from the sounding rocket 35.039 of the Visions-2 mission. The m-NLP system aboard rocket 35.039 consisted of four cylindrical Langmuir probes of length 39 mm and diameter .51 mm, biased to 3, 4.5, 6 and 7.5 V [39]. The model is trained with synthetic
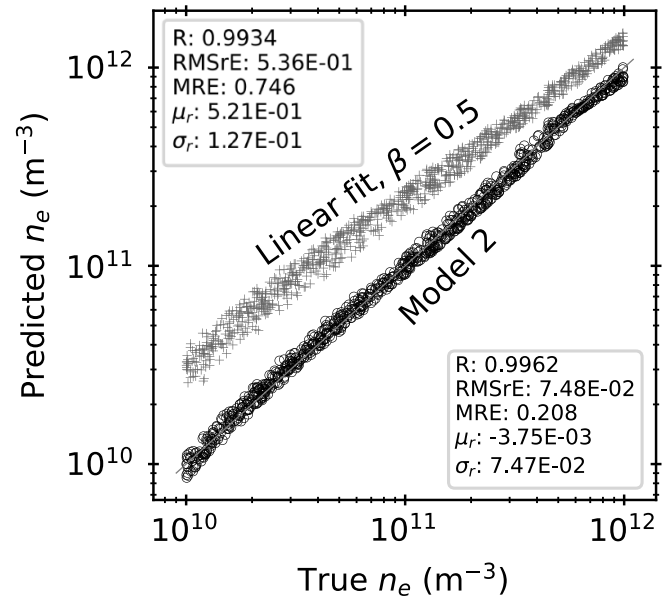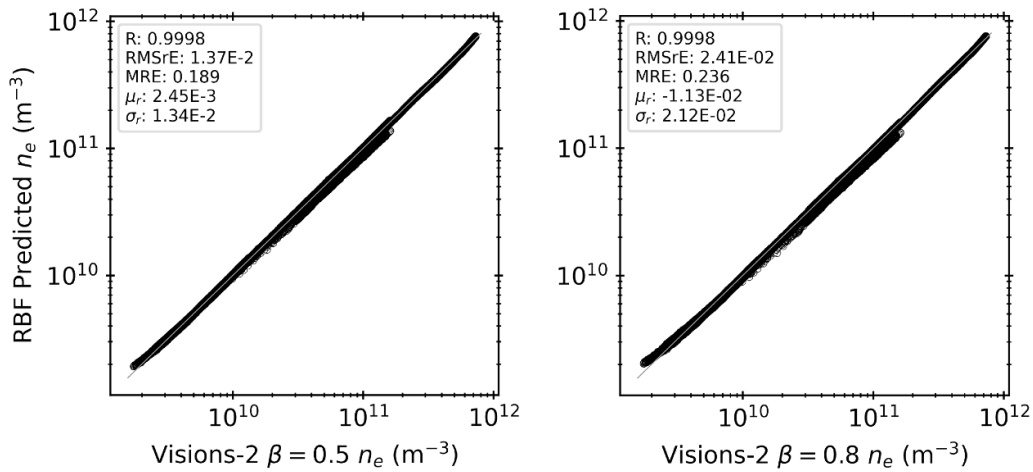

Figure 10. Correlation plot of the density obtained with model 2 (lower) with five pivots, applied to the `Langmuir` validation set. Predictions using a linear fit (upper) are included for comparison. Only a tenth of the points are shown.

data produced by the `Langmuir` library, and inferences of the density are compared with those obtained with two techniques used in this mission. The synthetic training data set of currents and densities was constructed for probes of diameter 0.51 mm, and length 40 mm. The same bias voltages of 3, 4.5, 6, and 7.5V were assumed as in the experiment. The data set was generated with randomly distributed electron densities in the range 10^{10}–10^{12} m^{-3} on a logarithmic scale and, using a linear scale, temperatures from 0.07 to 0.17 eV, and spacecraft floating potentials between -4 and -0.5 V. The RBF model was trained with 300 randomly selected currents and densities from a 10 000 node `Langmuir` data set, using five pivots. A comparison of inferred densities with those reported in the Visions-2 mission, as a function of time and altitude, is shown in figure 11. The two densities reported in the mission, shown in the figure, were obtained with Jacobsen's $\beta = 0.5$ linear fit, and the $\beta = 0.8$ non linear fit techniques.

As a final test of our proposed technique, RBF was trained with, and applied to experimental measurements. This is to ascertain whether the method is applicable to more general cases than those obtained with synthetic data constructed from simulations or computer models. To this end, the model was trained using five pivots and 300 randomly selected entries from the Visions-2 data set, consisting of 4-tuples of collected currents. In one case, the densities used in training and validation were inferred with Jacobsen's linear fit technique, while in the other case, they were inferred with a $\beta = 0.8$ nonlinear fit; both inferences being reported in the Visions-2 data set. A comparison between RBF inferences and experimental densities, including selected skill metrics, are shown in the two panels of figure 12 for the full Visions-2 data set. The ability of RBF to be trained with, and accurately reproduce densities inferred with these two different techniques, is yet
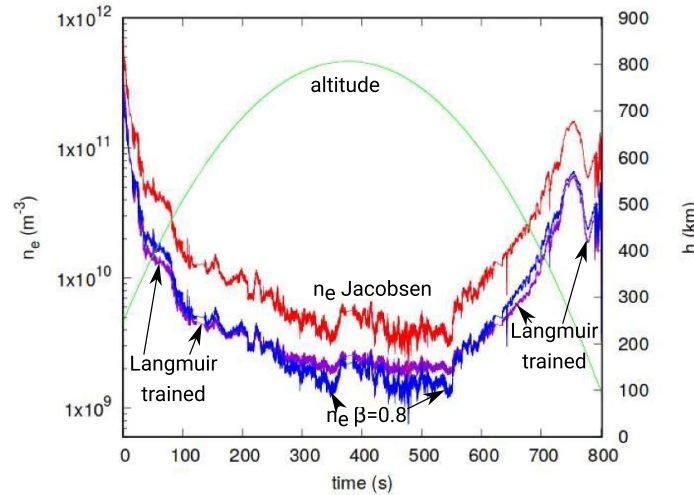
Figure 11. Comparison between densities from Visions-2 data inferred with Jacobsen's linear fit, $\beta = 0.8$ nonlinear fit, and RBF regression trained with data consisting of 300 randomly selected nodes from the `Langmuir` model, using five pivots.

Figure 12. Correlation plots of the density inferred with RBF trained with a 300 node subset of Visions-2 experimental data, against more than $4M$ experimentally inferred densities. Training and validation were made with densities inferred with the Jacobsen $\beta = 0.5$ linear fit (left), and the $\beta = 0.8$ nonlinear fit (right) technique.

another demonstration of the applicability of method 2 based exclusively on RBF regression. While the comparison made here cannot be used to ascertain the accuracy of either inference technique used in the experiment, it clearly shows that given accurately measured currents and densities, RBF can be used to construct high skill inference models for the density.

5. Summary and conclusion

New procedures are presented to infer a satellite floating potential and plasma density from currents collected with fixed-bias multi-needle Langmuir probes (m-NLP). The use of such probes was first considered by Jacobsen *et al* as a means of inferring plasma density, with high temporal and spatial resolution, independently of the electron temperature. Recognizing the limits of the OML approximation, with $\beta = 0.5$, linear and nonlinear least squares fit approaches have

been developed in order to infer the plasma density from probe measurements. In this paper we revisited these procedures and introduced two alternatives to infer plasma parameters from low level (L1B) m-NLP measurements. The first method makes use of a generalized orbital-motion-limited (OML) scaling law (see equation (1)), combined with radial basis function (RBF) regression to correct discrepancies obtained analytically. The second method relies solely on RBF regression to infer the density and floating potential. In both cases, physical parameters of interest are inferred from 4-tuples of currents collected by as many probes biased to known voltages. With the first method, the intermediate analytic expressions involve the plasma temperature, which can be obtained from independent measurements if possible, or from the regression procedure itself. In all cases considered, the inference of temperature comes with significant uncertainties, consistent with findings from previous studies. These large uncertainties result from the weak dependence of collected

currents on temperature which, as noted in previous studies, enables good quality inferences of the density, even with rough estimates of the temperature. With the second method, inference of both plasma density and satellite floating potential relies exclusively on RBF regression. The absence of analytic approximations in this case, implies that inferences are not biased to any *a priori* theory, and are therefore exclusively data-driven. The added challenge however, is that without being 'aided' by analytic approximations, regression now has to 'do all the work', resulting in possible accuracy loss for some parameters. The procedures reviewed and presented were assessed by applying them to a synthetic data set constructed with kinetic simulations, consisting of 4-tuples of collected currents with corresponding bias voltages, for a range of assumed densities, temperatures, and floating potentials. Although simulations do not account for the full complexity of processes at play near satellites in space, they do provide consistent data sets with known density, temperature and floating potential, from which inference algorithms can be tested. Model prediction skills were assessed graphically and quantitatively using adapted metrics. Consistently with what was reported by Hoang *et al* the linear fit approach leads to significant systematic overestimates of the density. We found that this can be corrected in part with RBF regression, to bring the centroid of predicted densities close to actual values. When inference is made using known (from the data sets) temperatures, all approaches provide good accuracy for the floating potential, but method 1 based on a combination of analytic scaling laws and regression is found to be appreciably more accurate. A loss of accuracy is noted in floating potentials resulting from method 1, and particularly from 4-parameter nonlinear fits, when temperatures are inferred from the model. Interestingly, this is not the case for predicted densities as, with model 1, inference accuracy is found to be slightly better when temperatures are calculated in the model. While the difference here is small (0.27 vs. 0.32 for the maximum relative error), this may result from the approximate nature of equation (1) assumed in method 1. Conversely, method 2 relying exclusively on RBF regression is found to have a relatively low inference skill for the floating potential, but an excellent one for the density. This indicates that an optimal strategy might require different algorithms to infer different physical parameters from a given instrument.

Finally, to show the applicability of direct RBF inference, to more general data sets, a proof-of-concept study was conducted using two additional and independent data sets. One was generated with the `Langmuir` code in which assumed plasma conditions and probe geometry were different from the ones assumed in our first assessments. The other consisted of currents measured experimentally in the Visions-2 mission, with densities inferred with two distinct inference algorithms. Although the true electron densities were unknown in this case, RBF produced results similar to those obtained when considering $\beta = 0.8$, which is believed to be more accurate than the original $\beta = 0.5$ method. Just as importantly, to show that RBF has the potential to accurately infer densities in experiments, given accurate training data (either through simulations or more accurate experimental methods), two RBF models were constructed by training on small subsets of the Visions-2 data with densities inferred from (a) Jacobsen's linear fit, and (b) the $\beta = 0.8$ nonlinear fit technique. It was then shown that RBF accurately reproduces densities, when trained with subsets of experimental data, independently of the experimental data analysis technique used to infer the density. These results provide strong evidence that direct RBF methods can be used to accurately infer densities from experimental data, given that the models are trained using sufficiently accurate data sets.

In conclusion, methods have been presented as promising to improve the accuracy of plasma density and satellite floating potentials inferred from m-NLP measurements. Our analysis shows that RBF alone (method 2) should be the preferred approach to infer densities from m-NLP measurements, whether training is made with synthetic simulation-based data or with data measured and validated experimentally. Conversely, based on our assessments made with data sets constructed with kinetic simulations, for which the empirical equation (1) is a good approximation, method 1 (combining regression and an empirical expression for the collected currents) should be preferred for inferring a satellite potential. More generally however, with different configurations of the probes relative to other satellite components, or plasma environment conditions, equation (1) might not accurately describe currents collected by the probes, which would then result in a loss of skill in inferences made with method 1. The possible variations on multivariate regression techniques and data sets are of course endless, and it would be possible to compare several more variants and data sets, which could of course lead to different results. While the determination of plasma density with m-NLP is not as straightforward as initially assumed on the basis of OML theory and the assumption of sufficiently long probes, this type of instrument offers interesting possibilities for measuring the density, as well a satellite floating potential. Part of our analysis is based on synthetic data generated with kinetic simulations, in which many processes at play in actual measurements are not accounted for. The results obtained are nonetheless sufficiently encouraging to motivate further computational and experimental studies with more physics, more detailed geometry, and broader expected space environment conditions, to support specific space missions.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iD

Joshua Guthrie ⬤ https://orcid.org/0000-0002-5160-5363

References

[1] Allen J E, Annaratone B M and de Angelis U 2000 On the orbital motion limited theory for a small body at floating potential in a Maxwellian plasma *J. Plasma Phys.* **63** 299–309

[2] Barjatya A and Merritt W 2018 Error analysis of multi-needle Langmuir probe measurement technique *Rev. Sci. Instrum.* **89** 043507

[3] Barjatya A, Swenson C M, Thompson D C and Wright K H Jr 2009 Invited article: data analysis of the floating potential measurement unit aboard the international space station *Rev. Sci. Instrum.* **80** 041301

[4] Bettinger R T and Anthony Chen A 1968 An end effect associated with cylindrical Langmuir probes moving at satellite velocities *J. Geophys. Res.* **73** 2513–28

[5] Bilitza D, Altadill D, Truhlik V, Shubin V, Galkin I, Reinisch B and Huang X 2017 International Reference Ionosphere 2016: from ionospheric climate to real-time weather predictions *Space Weather* **15** 418–29

[6] Buhmann M D 2003 *Radial Basis Functions: Theory and Implementations* vol 12 (Cambridge: Cambridge University Press)

[7] Chalaturnyk J and Marchand R 2019 A first assessment of a regression-based interpretation of Langmuir probe measurements *Front. Phys.* **7** 63

[8] Deca J, Lapenta G, Marchand R and Markidis S 2013 Spacecraft charging analysis with the implicit particle-in-cell code iPic3D *Phys. Plasmas* **20** 102902

[9] Feoktistov V 2006 *Differential Evolution* (Berlin: Springer)

[10] Fisher L E *et al* 2016 Including sheath effects in the interpretation of planar retarding potential analyzer's low-energy ion data *Rev. Sci. Instrum.* **87** 043504

[11] Folkestad K, Hagfors T and Westerlund S 1983 Eiscat: an updated description of technical characteristics and operational capabilities *Radio Sci.* **18** 867–79

[12] Gill E, Sundaramoorthy P, Bouwmeester J, Zandbergen B and Reinhard R 2013 Formation flying within a constellation of nano-satellites: the QB50 mission *Acta Astronaut.* **82** 110–7

[13] Godyak V A and Alexandrovich B M 2015 Comparative analyses of plasma probe diagnostics techniques *J. Appl. Phys.* **118** 233302

[14] Hoang H *et al* 2018 The multi-needle Langmuir probe system on board NorSat-1 *Space Sci. Rev.* **214** 75

[15] Hoang H, Røed K, Bekkeng T A, Moen J I, Spicher A, Clausen L B N, Miloch W J, Trondsen E and Pedersen A 2018 A study of data analysis techniques for the multi-needle Langmuir probe *Meas. Sci. Technol.* **29** 065906

[16] Hoang H, Røed K, Bekkeng T A, Moen J I, Spicher A, Clausen L B N, Miloch W J, Trondsen E and Pedersen A 2018 A study of data analysis techniques for the multi-needle Langmuir probe *Meas. Sci. Technol.* **29** 065906

[17] Hoang H *et al* 2019 The multi-needle Langmuir probe instrument for QB50 mission: case studies of ex-alta 1 and hoopoe satellites *Space Sci. Rev.* **215** 1–19

[18] Gray J B and Ling R F 1984 K-clustering as a detection tool for influential subsets in regression *Technometrics* **26** 305–18

[19] Hoskinson A R and Hershkowitz N 2006 Effect of finite length on the current–voltage characteristic of a cylindrical Langmuir probe in a multidipole plasma chamber *Plasma Sources Sci. Technol.* **15** 85

[20] Ishikawa T and Matsunami M 1997 An optimization method based on radial basis function *IEEE Trans. Magn.* **33** 1868–71

[21] Jacobsen K S, Pedersen A, Moen J I and Bekkeng T A 2010 A new Langmuir probe concept for rapid sampling of space plasma electron density *Meas. Sci. Technol.* **21** 085902

[22] Jaeger E F, Berry L A and Batchelor D B 1991 Characteristics of a cylindrical Langmuir probe of finite length *J. Appl. Phys.* **69** 6918–22

[23] Johanning D, Seifert W and Best A 1985 Analytical density correction for cylindrical Langmuir probes showing end effects *Plasma Phys. Control. Fusion* **27** 159

[24] Laframboise J G 1966 Theory of spherical and cylindrical Langmuir probes in a collisionless, Maxwellian plasma at rest *PhD Thesis* University of Toronto

[25] Langmuir I and Mott-Smith H M 1926 The theory of collectors in gaseous discharges *Phys. Rev.* **28** 99

[26] Lei J, Liu L, Wan W and Zhang S-R 2005 Variations of electron density based on long-term incoherent scatter radar and ionosonde measurements over Millstone Hill *Radio Sci.* **40** 2

[27] Lomidze L, Knudsen D J, Burchill J, Kouznetsov A and Buchert S C 2018 Calibration and validation of swarm plasma densities and electron temperatures using ground-based radars and satellite radio occultation measurements *Radio Sci.* **53** 15–36

[28] Marchand R 2012 PTetra a tool to simulate low orbit satellite-plasma interaction *IEEE Trans. Plasma Sci.* **40** 217–29

[29] Marchand R and Resendiz Lira P A 2017 Kinetic simulation of spacecraft–environment interaction *IEEE Trans. Plasma Sci.* **45** 535–54

[30] Marchand R *et al* 2014 Cross-comparison of spacecraft-environment interaction model predictions applied to solar probe plus near perihelion *Phys. Plasmas* **21** 062901

[31] Marholm S, Marchand R, Darian D, Miloch W J and Mortensen M 2019 Impact of miniaturized fixed-bias multineedle Langmuir probes on cubesats *IEEE Trans. Plasma Sci.* **47** 3658–66

[32] Marholm S and Darian D 2019 langmuirproject/langmuir: 0.6.0 (Version 0.6.0) *Zenodo* (https://doi.org/10.5281/zenodo.3464277)

[33] Marholm S and Marchand R 2020 Finite-length effects on cylindrical Langmuir probes *Phys. Rev. Res.* **2** 023016

[34] Murphy K and Murphy T B 2019 Gaussian parsimonious clustering models with covariates and a noise component *Adv. Data Anal. Classif.* **14** 293–325

[35] Rusu C and Rusu V 2006 Radial basis functions versus geostatistics in spatial interpolations *IFIP Int. Conf. on Artificial Intelligence in Theory and Practice* (Springer) pp 119–28

[36] Saad Y 2003 *Iterative Methods for Sparse Linear Systems* 2nd edn (Philadelpha, PA: SIAM)

[37] Shakhmatov E, Belokonov I, Timbai I, Ustiugov E, Nikitin A and Shafran S 2015 SSAU project of the nanosatellite SamSat-QB50 for monitoring the earth's thermosphere parameters *Proc. Eng.* **104** 139–46

[38] Shukla X U and Parmar D J 2016 Python–a comprehensive yet free programming language for statisticians *J. Stat. Manage. Syst.* **19** 277–84

[39] Spicher A, David Michael Bang-Hauge and University of Oslo (Norway) Espen Trondsen. m-nlp on visions-2 35.039 [dataset]. norstore. (https://doi.org/10.11582/2021.00004)

[40] Storey L R O 1998 What's wrong with space plasma metrology? *Geophys. Monogr.-Am. Geophys. Union* **102** 17–22

[41] Sudit I D and Claude Woods R 1994 A study of the accuracy of various Langmuir probe theories *J. Appl. Phys.* **76** 4488–98

[42] Tuszewski M and Tobin J A 1996 The accuracy of Langmuir probe ion density measurements in low-frequency RF discharges *Plasma Sources Sci. Technol.* **5** 640