Referee: 1  
  
 COMMENTS TO THE AUTHOR(S)  
 [Preliminary assessment - proceed with review]  
 xxx  
  
 Referee: 2  
  
 COMMENTS TO THE AUTHOR(S)  
 The manuscript introduces a novel approach to the characterization of low-temperature plasma. Although the findings are promising, certain aspects of the methodology and results warrant additional check prior to publication.

**Comment 1:** Abstract  
 The abstract of the paper introduces a novel technique that could offer significant insights for LTP applications.  
 However, its presentation of the methodology and results requires improvement for clarity and impact.  
 Firstly, the description of the methodology is overly general. It is not immediately evident how the methodology is applied to Low-Temperature Plasmas (LTPs) or how it operates within this context. A more specific and detailed explanation is needed to highlight the innovative aspects of the approach. The abstract should clarify how the methodology contributes to solving key challenges in plasma systems and specify its application in LTP environments.

We have included the innovative aspects of the approach, particularly the adapted L1 penalty method. This is in line 10 of the abstract.

**Comment 2:** Secondly, the outcomes of the research should be more prominently featured. At present, the results are either vaguely described or not sufficiently emphasized. A more explicit presentation of the findings would significantly enhance the abstract's effectiveness and help readers better grasp the significance of the study in advancing the field of LTPs.

We have also added that we found that the dimensionless electron density and electron temperature were directly a function of reaction parameters for the system under study. This is in line 16 of the abstract.  
 **Comment 3:** Line 25 Page 1: Scaling and Similarity Relations  
 The use of scaling and similarity relations in LTP systems requires further clarification. Specifically, how do these concepts aid in the design and optimization of LTP systems? It would be helpful to provide briefly concrete examples of how similarity principles have been applied in plasma research to better illustrate their relevance.

We have noted that similarity parameters enable us to directly scale-up (systems with the same similarity parameter have the same behavior), and scaling relations can enable trade-offs in design such as the impact of current on joule heating. This is in line 5 of the abstract.

**Comment 4:** Line 29 Page 1: Sparse Dimensionless Numbers  
 The term "sparse dimensionless numbers" should be clearly defined, and I suggest providing some examples for clarity.  
We have added this definition.

**Comment 5:** Line 39 Page 1: Physical Insights … Scaling Laws  
 The statement about the physical insights provided by interpretable dimensionless relations and scaling laws should specify which plasma characteristics are being systematically analyzed. This clarification would offer readers a clearer understanding of the scope and potential applications of the research.  
  
This was a statement meant to highlight that there are broad applications for the method used. We believe our method can be used to discover similarity and scaling relations for a range of discharges and chemistries. We have additionally added specifically which plasma characteristics are under study in line 16.

**Comment 6:** Text  
 The introduction is well-structured and offers a solid foundation for the study. However, a key concern arises regarding the application of the Buckingham Pi methodology. This technique is typically used for independent variables, yet plasma measurement data, as referenced in the literature, may not be entirely independent. It would be helpful to address this issue more explicitly and explain how interdependencies between plasma parameters are handled in this context.

The Buckingham-Pi theorem guarantees that a physical law can be represented in any unit system, and the physical law can even be represented as independent dimensionless numbers. So, we enforce that the target dimensionless number Pi\_0 and the fitted Pi\_i are not drawn from the same sets of variables. To use Section 3.1 as an example, the target dimensionless number Pi\_0 is a function of n\_e; n\_e is withheld from the candidate process variables that can be used to construct Pi\_1 in Table 2 to avoid this problem. To further avoid problems regarding potential dependence in the process variables used to construct Pi\_1, we use our l1 norm to penalize the size of Pi\_1. The value in this is that if two variables carry the same information (because they are not independent), the system will discard one of them as redundant.

**Comment 7:**  
 Furthermore, the development of a global model is mentioned, but the rationale for its creation is unclear. The authors suggest that the model is used to calculate plasma parameters that are not directly measurable, but the process of integrating physical measurements with approximate solutions in a predictive model needs further clarification. How are two variables of different nature—physical measurements and model-derived solutions—combined in the analysis?  
  
In this work, we aim to show how our method of learning dimensionless relationships could be applied to a LTP system. In this case, we use a LTP model to produce values we could not measure. For the sake of demonstration, we then act as if both sets of values (calculated and measured) represented the ground truth. We then discover dimensionless relationships among the data. A clarifying statement has been added to make this more explicit on page 8.

**Comment 8:** Variable Selection and Exclusion  
 The paper frequently refers to the selection and exclusion of certain variables. While it is understood that these decisions are based on criteria such as the order of magnitude of reaction rates and Buckingham Pi algorithms, the methodology behind these choices is not sufficiently explained. Initially, readers may assume that variables derived directly from experimental data will be used, but it becomes apparent that variables from theoretical models are also introduced. The criteria for including these model-derived variables should be more thoroughly explained. For instance, why are certain variables from the theoretical model necessary? What would happen if they were excluded? How might this affect the predictions made by the Buckingham Pi methodology?

Why model-derived variables, and why are they necessary?

These additional variables let us capture and relate more phenomenon such as the electron density. Treating variables from this model as the ground-truth enables us to interpret the results of our dimensionless-discovery method. In particular, we can compare the position and form of variables in the discovered relationship to rates. Further, using variables from the same model that predicts the electron density and on-pulse electron temperature ensures that the variables we use are physically related to the electron density and on-pulse electron temperature. This is noted on page 8.

Why are only some of the variables included? First, it is not always possible to get the method to converge when all variables are included. When our discovery method does converge when including all variables, it is the same as the case of including the top six rates. Second, as in Figure 6, we show that as we increase the number of variables considered, that the performance of the discovered relationship improves. We analyze why. We show what happens if many are excluded. This is noted in the first paragraph of Sec 3.1.

If no variables from the model were included then it would be more challenging to interpret the final result as a ratio of rates based on the model, and hence to analyze how useful our presented method might be. ~~If no variables from the model are included, then it is likely that there would not be sufficient information to predict the dimensionless electron density, as it is derived from the model. It is possible that the measured variables could be used to predict the dimensionless electron temperature, these measured variables being on-pulse time, frequency, volume, area, length, voltage, charge during the on-pulse, gas temperature (constant).~~

**Comment 9:** Section 3.1  
 While the explanation of the results is generally well done, further clarification is needed on why dimensionless numbers are of interest, especially when they can be calculated from the global model. The outputs of the Buckingham Pi model are not entirely clear, and it would be helpful to explicitly state how these dimensionless numbers can be applied in plasma applications.

We are grateful for the kind words. To address your concerns:

We have elaborated on why these particular target dimensionless numbers (dimensionless electron density and dimensionless electron temperature) are useful on page 4; this in addition to existing explanations on the general utility of dimensionless numbers on the same page.

We seek to demonstrate the accuracy of our method, which may have great utility when a model is not present, or it is difficult to draw conclusions from the existing model. For the case where no model is present, only measurements, it may be possible to use this method to discover dimensionless relationships from the data. When compared to the existing model, these learned dimensionless relationships are much more understandable yet still accurate. For even more complex models or simulations like COMSOL, such simplified representations can be very useful.

Finally, we briefly discuss in Section 3.1 (paragraph 2, page 18) that the discovered relationship for electron density implies that any set of He reactors that follow our assumptions will have the same electron density for the same reaction parameters. We have added some additional utility: in a different system where the products were of interest, it might be very useful to know that engineering a higher dimensionless electron density might change the rate of these two reactions and associated products, which could be valuable from a process engineering perspective. We have a similar discussion at the ends of Section 3.2 and 3.3.

**Comment 10:** Line 6 Page 2: Manipulable Process Parameters  
 The phrase "manipulable process parameters" requires clarification. Which specific parameters are being referred to, and how are they manipulated within the context of plasma experiments?  
  
Any process parameter that can be changed directly by the scientists, such as applied voltage or electrode length. Quantities like electron density cannot be directly manipulated. We have added a definition for this term where it appears on page 2.

**Comment 11:** Line 15 Page 2, Line 6 Page 3: similarity relationships, similarity parameters  
 The reference to "similarity relationships" in this context remains vague. Could the authors specify the exact relationships being referred to and how these relationships help scale plasma discharges for practical applications?

We have replaced “similarity relationship” with similarity. This is because systems with the same similarity parameters have the same behavior across scales and hence are said to be similar (“Scaling”, by Barenblatt).

We appreciate this recommendation and have added examples so that it is more concrete to the reader on page 3. We note that similarity parameters have been very successful in scaling other systems such as pipe flow and airplane design, and there is opportunity to apply this to plasma systems. We have provided examples of similarity parameters in plasma science and in scaling plasma systems. We have included preventing accidental discharges using the Paschen curve or microdischarges, scaling up dielectric barrier discharges at vacuum, sprites in the atmosphere, the townsend-streamer transition, and other plasma phases.

**Comment 12:** Line 22 Page 2: Generalized Homogeneity  
 The concept of generalized homogeneity is introduced, but the implications for plasma systems are not entirely clear. How does this property apply when the geometry of the plasma reactor changes? Plasma is inherently non-homogeneous, so it would be useful to elaborate on how generalized homogeneity can still hold in cases where plasma is not uniform, particularly in Low-Temperature Plasma (LTP) systems.

Generalized homogeneity does not refer to the homogeneity of the discharge, but rather is a physical property that indicates that physics-based relationships are true regardless of the units of measurement. On review, we find that this term clutters the discussion and has been removed.

**Comment 13:** Line 51 Page 2: Dimensional Analysis…  
 The authors mention that dimensional analysis has been applied to fully ionized plasmas and magnetohydrodynamics (MHD). Given that LTPs are often non-homogeneous and produced under atmospheric conditions, it would be beneficial to clarify whether the Buckingham Pi methodology is applicable in such cases. Plasma in these conditions typically involves non-homogeneous phenomena, and the interdependence of process variables complicates the analysis.

Homogeneity: applied to nonhomogeneous systems or systems with sharp changes in the physics, such as shockwaves (solitons) and boundary layers (“Scaling”, by Barenblatt). It has also successfully been applied to analyze inhomogeneous systems through methods such as bulk averaging or the microelement method (“Scaling Analysis in Modeling Transport and Reaction Processes”, by Krantz), as seen in the case of pressure drop in a packed conduit (Ergun equation) We have added this to page 2.

~~Interdependent pi groups: interdependent pi groups are pi groups that are products of other pi groups. These pi groups are redundant. We control for this by choosing the number of dimensionless groups considered, which enables the user to start with one pi groups (where there can be no redundant pi groups) and expand from there. If the user has included too many pi groups such that redundant or dependent pi groups are included, our system can account for this by penalizing the size of dimensionless groups. In this case, if a user selected two pi groups when only one is required, the L1 penalty used in (13) should encourage the algorithm to eliminate one of the pi groups.~~

~~Interdependent variables: if two variables “a” and “b” are interdependent and we attempt to fit y = f(a,b,c), the l1 penalty should discard one of “a” and “b” in fitting to y. However, it is possible that two or more interdependent variables may be used when only one is needed due to the dimensionless constraint, where these otherwise superfluous variables are included to cancel the dimensions in a pi group.~~

**Comment 14:** Line 42 Page 2, Line 51 Page 3: ..Scarce Data…  
 The distinction between "sparse" and "scarce" data is not clearly explained. Could the authors clarify the relationship between these terms, and how they apply to plasma processes?  
  
We clarify both here. In the manuscript, we clarify the meaning of “sparse” in this context in the same location, page 4.

We use “scarce” to indicate that only a small quantity of data is needed. This contrasts our method with other machine learning methods that find lower dimensional representations of data, such as deep autoencoders, which can be very data hungry. We believe “scarce” is clear in context so long as “sparse” is explained.

We reserve “sparse” to indicate a regression model where superfluous independent variables or basis functions have been discarded. This is judged by the goodness-of-fit, that is, if discarding a variable only marginally impacts the goodness-of-fit, then it is likely that this variable is not meaningful. This is discussed and examined in Section 3, such as in Figures 7 and 8. Similarly for dimensionless numbers, this means that the dimensionless number has only a small number of terms.

**Comment 15:** Line 49 Page 3: Sparse Nondimensional Relations  
 The concept of sparse nondimensional relations, while introduced, is not clearly explained in relation to plasma parameters. It would be helpful to provide more context for how the term "sparse" applies to physical plasma parameters, and how the proposed methodology is suited for handling these relationships. Without this explanation, the relevance of the approach may be unclear to the reader.  
  
We are thankful for this recommendation. We have clarified what is meant by sparse nondimensional relations, namely, dimensionless equations that have dimensionless groups with only the minimum necessary terms, and simple relationships between the dimensionless groups. We have further connected this to plasma and why it may be of interest. This has been added in red at the top of page 4.

**Comment 16:** Line 50 Page 6: Constant k1  
 The constant "k1" is mentioned but not defined. Could the authors provide more details about this constant?

**I will add this, it’s just a constant of fitting. Update: following the citation chain to get a better descripion than “an unimportant constant”, which is what our source says.**  
  
 **Comment 17:** Line 41 Page 7: …Global Plasma Model…  
 The use of a global model alongside the Buckingham Pi methodology raises some questions. Specifically, how can data of such different natures—experimental measurements and model-derived data—be combined effectively? Given that global model data may be subject to debate regarding its accuracy, it is important to clarify how discrepancies between models and experimental data are addressed to ensure the reliability of the analysis.  
  
We take physical measurements and then feed them into the global model, which is taken to be the ground truth. With this ground truth, we can then apply our data-driven discovery of dimensionless numbers and nondimensional relationships. This has been made explicit on page 7, Section 2.2.

**Comment 18:** Line 31 Page 13: Here, we aim to learn a sparse model for predicting…  
 Why use the Buckingham Pi model when the number density has already been calculated from the global model? Clarifying this would help the reader understand the necessity of employing both approaches.

The purpose of this work is to demonstrate that this method can extract simple and interpretable relationships from scarce data, where that scarce data is derived from complex systems. In this case, we use the global model as the ground truth to produce plasma data. We then use our data-driven discovery method to find relationships in the data. The validity of these relationships can be compared to the model. This is recorded in Section 2.2 **(should we add this to the preamble to Section 3, to make it clearer for the reviewer?).**

**Comment 19:** Line 39 Page 15: Discarding Unnecessary Process Variables  
 The statement regarding the automatic discarding of unnecessary process variables raises the question of how these variables are assessed from a physical perspective, especially in the context of plasma processes. What criteria are used to determine the relevance of these variables, beyond computational considerations?  
  
We are demonstrating that our method for data-driven discovery of nondimensional relations can find a nondimensional relationship from a set of values known to be related. In this case, we generate our target values (the dimensionless electron temperature and dimensionless electron density) from a model that is treated as the ground truth. We can hence guarantee that there is some relationship between the process variables used and the target dimensionless numbers. We can then determine which process variables are most relevant by their ability to accurately predict the target dimensionless numbers (for example, Figures 6 and 7) and even extrapolate to unseen data (for example, Figure 8). Figure 7 is particularly important, as it can be seen that adding more variables does not improve the predictive capability of the discovered relationship. This implies that there is a tight link between the target dimensionless number and the kept process variables. The exact form and meaning of that physical relationship is discussed.

**(Our method aims to find a dimensionless number or numbers that can predict the target dimensionless number accurately. We ensure our process variables are relevant by drawing them from the energy and particle balance. It is likely that the other process variables play an indirect role, for example in influencing the electron temperature and hence the kinetic parameters. We also remind the Reviewer that the “computational considerations” are considerable: as in Figure 6, many of the process variables cannot predict the target dimensionless number with any significant accuracy. When the retained process variables are added, it can be seen in Figure 6 that the size of the dimensionless number collapses—because the other process variables are not nearly so predictive in the range explored. To more certainly state that only our retained process variables matter, we could do a follow-up study by expanding the range of conditions considered).**

**Comment 20:** Line 30 Page 17: explicitly dependent on operating parameters…  
 The term "explicitly" in relation to dependence on operating parameters requires further clarification. Why is this explicit dependence important in the context of plasma applications?  
In this case, an explicit relationship means that operating parameters such as voltage or volume would appear on the RHS of (16).

An explicit relationship would be very informative and useful for engineering efforts. For example, if we found that the dimensionless electron density had some power law dependence on voltage or volume, we could then recognize that we could tune the electron density directly by changing these values. We expect voltage and volume to still matter, but that their influence is expressed by changing these reaction parameters.

We found that the dimensionless electron density and dimensionless electron temperature were a function of reaction parameters, indicating a different reactor would behave identically for the same reaction parameters. This is very similar to what was found by Bell in analyzing a dielectric barrier discharge (doi: 10.1021/i160033a026), finding that the reactor could be scaled up provided the reaction parameters (which were functions of the reduced electric field in his work) were the same.  
  
 Referee: 3  
  
 COMMENTS TO THE AUTHOR(S)  
 The manuscript titled "Data-Driven Discovery of Sparse Dimensionless Numbers and Nondimensional Relations for Low-Temperature Plasmas" introduces a learning-based data-driven method for discovering the expressions of dimensionless numbers in LTPs. The authors use helium gas in the ns-DBD reactor to generate plasma for data collection and experimental validation. The expressions for the dimensionless numbers ne/ng and Te/Tg are discovered, and the practical significance of each variable in the expressions within the context of plasma reactions is discussed. The following comments are suggested to improve the manuscript.

**Comment 1:** 1. The reason for studying the target dimensionless numbers ne/ng and Te/Tg should be given.

We appreciate this comment and have added our rationale to make for a clearer read. We have added this to page 4. The reason is that both have been used to understand the departure from equilibrium, and both are important to reaction kinetics in this system.

**Comment 2:** 2. "LTP" and "LTPs" are used in the manuscript. Besides, "low-temperature plasmas" is used in Abstract and in Introduction, but "low-temperature plasma" is used in Keywords. Please standardize the notation to one consistent form.

Thank you for the recommendation. We have chosen low-temperature plasmas and LTPs.

**(I am confused by this. I started making corrections, but then stopped. Does the reviewer mean to stick with singular vs plural, or with the acronym vs written out?)**

**Comment 3:** 3. Figure 1 is not intuitive enough. The photograph of the reactor should be given.

Thank you. We have added a photo of the reactor, complete with the OES fiber and light shield.

**Comment 4:** 4. From the flowchart in Figure 5, it can be seen that the loss function used in the model is in the simple form of equation (9). Is it necessary to derive the more complex form of the loss function as shown in equation (13)?

Yes. (13) is the most general form. In this way, other scientists can borrow our equations and our code if they wish to repurpose it for their own ends.

**Comment 5:** 5. "Number of Largest Rates" axis in Figure 6 and in Figure 9 need to be further explained.

We are grateful for the recommendation. We have elaborated on page 15.

**Comment 6:** 6. The relationship between q ≤ b dimensionless numbers is represented by equation (2), but no evidence is provided. Please explain the purpose and the feasibility.

We have clarified this at the start of Section 2.3.  
In brief, a physical relation among “p” variables with “d” independent physical dimensions can be made into a nondimensional relationship by rearranging it into p-d = b dimensionless grouping of variables, called dimensionless numbers. The reason that there are p-d =b possible independent groups springs from the rank-nullity theorem.  
In our case, we supply “p” many process variables but do not know which of the “p” many process variables are relevant. If only “v” many process variables are relevant for v<p, then we would need v-d =q dimensionless numbers, where q<b. This leads to equation (2), which is now (4)..

**Comment 7:** 7. The manuscript mentions "Π0 denotes a target, user-specified dimensionless number that we look to predict" and "Πi are q−1 dimensionless numbers that are constructed from process variables x" in page 9. Why can Πi (the input dimensionless number) be constructed from process variables x, while Π0 (the target dimensionless number) cannot and needs to be derived from equation (2)? Please explain the difference between input dimensionless numbers and target dimensionless number, and how they are defined.

Π\_0 (the target dimensionless number) is not derived from equations given, but is instead supplied by the user.

It is theoretically possible to derive all dimensionless numbers from the process variables. However, that is a serious challenge from an algorithmic perspective, and suitable for future work. In the current work, the dimensionless numbers Π\_i are found by determining what dimensionless numbers best predict the target dimensionless numbers. Designing an algorithm that can suggest both by Π\_0 and Π\_i is a vastly larger problem.

**Comment 8:** 8. The manuscript mentions "However, since the linear system of equations (4) that must be solved to find w is underdetermined (i.e., d < p), an infinite number of solutions can exist for w." in page 11. But there is no evidence to prove that d < p.

This is the problem of practical interest. Recall that “d” is the number of dimensions, and “p” is the number of process variables supplied by the user. Recall additionally that **D**, the real-valued dimensions matrix, is size dxp. As noted in the manuscript, “d” cannot exceed 7; in this work, we did not exceed 5. A scientist who has many process variables has a large “p”; for d<p, we are guaranteed that the nullity of **D** is not an empty set by the rank-nullity theorem. For the case considered in this work wherein the scientist does not know which process variables are predictive of the target dimensionless number, this can lead to p much larger than d. For p much larger d, the dimensionality of the nullity of D can grow very quickly, making it very difficult to determine what process variables or dimensionless numbers matter.   
  
This has been specified in page 3 on the discussion of previous work on data-driven discovery of similarity and scaling parameters. It has been additionally been added to the beginning of page 10, Sec. 2.3.

**Comment 9:** 9. In equation (9), the variable γi is a regularization term, as the manuscript mentions " where γi denote the coefficients of the basis vectors" in page 11. Please explain why basis vectors can be sparsified and what is the influence of discarding some basic vectors on dimensionless numbers.

We want to be clear that we are sparsifying the coefficients of basis vectors. In (8), we see that the exponents that make up the dimensionless number are related to the basis vectors. That is, the exponents are a linear combination of the basis vectors. By sparsifying the exponents vector **γ\_**i, we are discarding some of the basis vectors by multiplying them by zero.

We can achieve this by penalizing the l1 norm of **γ\_**i seen in (9), now (11) in the updated manuscript. The reason this can work is that the l1 norm of a vector is the sum of the absolute of each term in the vector; shrinking this can only be achieved by shrinking the individual entries of **γ\_**i. The entries correspond to coefficients of basis vectors. A lot of these entries are preferentially set to zero (show image form LASSO) due to the geometry of the fitting problem.

What is the influence of discarding some basis vectors on dimensionless numbers? The answer is that we remove some of the basis vectors, which may shrink the exponents of the resulting dimensionless number. By setting a variable to have an exponent of zero, this variable is effectively “discarded” from the dimensionless number. To ensure we set as many of the exponents to zero, we modify our loss function as in (14) and (15) with the intent that we penalize the largest exponent.