



OPEN ACCESS

EDITED BY

David Benoit,
University of Hull, United Kingdom

REVIEWED BY

Kotomi Taniguchi,
National Astronomical Observatory of Japan (NAOJ), Japan
Francesco Fontani,
Osservatorio Astrofisico di Arcetri (INAF), Italy

*CORRESPONDENCE

Jacobo Aguirre,
✉ jaguirre@cab.inta-csic.es

RECEIVED 03 February 2025

ACCEPTED 10 July 2025

PUBLISHED 30 July 2025

CITATION

Fernández-Ruz M, Jiménez-Serra I, Castro M, Ruiz-Bermejo M and Aguirre J (2025) Strong parameter hierarchy in the interstellar phosphorus chemical network.
Front. Astron. Space Sci. 12:1570525.
doi: 10.3389/fspas.2025.1570525

COPYRIGHT

© 2025 Fernández-Ruz, Jiménez-Serra, Castro, Ruiz-Bermejo and Aguirre. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Strong parameter hierarchy in the interstellar phosphorus chemical network

Marina Fernández-Ruz^{1,2}, Izaskun Jiménez-Serra¹,
Mario Castro^{2,3}, Marta Ruiz-Bermejo¹ and Jacobo Aguirre^{1,2*}

¹Centro de Astrobiología (CAB), CSIC-INTA, Madrid, Spain, ²Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain, ³Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid, Spain

Phosphorus-bearing molecules are fundamental for life on Earth, yet their astrochemical origins remain poorly understood. Their formation in the interstellar medium has been challenging to elucidate due to limited spectroscopic detections and the reliance on theoretical models that depend on numerous kinetic parameters whose values are very uncertain. Multi-parameter models often suffer from “sloppiness”, where many parameter combinations exhibit negligible influence on model outcomes, while a few dominate system behavior. In this study, we introduce the Fisher Information Spectral Reduction (FISR) algorithm, a novel and computationally efficient method to reduce the complexity of such sloppy models. Our approach exposes the strong parameter hierarchy governing these systems by identifying and eliminating parameters associated with insensitive directions in the parameter space. Applying this methodology to the phosphorus astrochemistry network, we reduce the number of reaction rate coefficients from 14 to 3, pinpointing the key reactions and kinetic parameters responsible for forming PO and PN, the main phosphorus-bearing molecules typically detected in interstellar space. The simplified model retains its predictive accuracy, offering deeper insights into the mechanisms driving phosphorus chemistry in the interstellar medium. This methodology is applicable to multi-parameter models of any kind and, specifically in astrochemistry, facilitates the development of simpler, more realistic and interpretable models to effectively guide targeted observational efforts.

KEYWORDS

phosphorus astrochemistry, interstellar medium, astrobiology, dynamical systems, chemical reaction networks, kinetic parameters, complexity reduction, parameter sensitivity

1 Introduction

Phosphorus (P) is an element of significant astrobiological importance due to its abundance in biomass and its critical role in essential biochemical functions: the backbone of nucleic acids contains sugar-phosphates, phosphorylated molecules act as energy carriers, and cellular membranes contain phospholipid head groups (Walton et al., 2023). Interestingly, despite its ubiquity in life on Earth, phosphorus is far less abundant on cosmic scales than other essential elements for life, such as hydrogen (H), carbon (C), oxygen (O), and nitrogen (N)—a phenomenon often referred to as the phosphorus enigma. In fact, phosphorus ranks 18th in cosmic abundance, orders of magnitude lower than

other biogenic elements (Maciá-Barber, 2020). With regard to the origin of terrestrial phosphorus, it is worth noting that the chemical complexity of early Earth was augmented by the influx of molecular compounds inherited from its parental molecular cloud during the process of planetary formation and the subsequent impact of asteroids and comets on its surface. Interplanetary dust, meteorites, and large impactors may have deposited reactive, reduced oxidation state phosphorus onto early Earth's surface. This process, supplemented by terrestrial P-reduction pathways, enriched prebiotic environments with reactive P, including organic phosphonates with carbon-phosphorus bonds, as identified in the Murchison meteorite (Cooper et al., 1992). Furthermore, the recent detection of PO in comet 67P/Churyumov-Gerasimenko (Rivilla et al., 2020) and of phosphorus-rich grains in asteroid Ryugu samples (Pilorget et al., 2024), indicates that minor Solar System bodies can act as reservoirs for phosphorus-bearing compounds. Therefore, the detection of phosphorus sources in the interstellar medium (ISM) is key to trace the evolution of phosphorus during the Solar System formation and unveil how this element became available for the origin of life.

In recent decades, the interstellar chemistry of phosphorus has emerged as a promising area of research driven by high-sensitivity instrumentation such as the Atacama Large Millimeter/submillimeter Array (ALMA). To date, more than 300 molecules have been detected in interstellar and planetary environments, including complex organic molecules (COMs) with potential prebiotic significance. Despite this impressive detection capacity, only seven P-bearing molecules have been unambiguously identified in the ISM: PO (Tenenbaum et al., 2007; Rivilla et al., 2016), PO⁺ (Rivilla et al., 2022), PN (Turner and Bally, 1987; Ziurys, 1987; Fontani et al., 2016), CP (Guelin et al., 1990), HCP (Agúndez et al., 2007), CCP (Halfen et al., 2008), and PH₃ (Agúndez et al., 2008; Tenenbaum and Ziurys, 2008). Among these, only PO, PO⁺, and PN have been detected in star-forming regions, but with low abundances (the detected abundances of PO and PN relative to H are on the order of 10⁻¹⁰ and 10⁻¹¹, respectively; see e.g., Lefloch et al. (2016), Rivilla et al. (2018), Rivilla et al. (2020), Bernal et al. (2021), Fontani et al. (2024), Lefloch et al. (2024), Scibelli et al. (2025)). In addition, astronomical observations show that PO is systematically more abundant than PN, with abundance ratios of [PO]/[PN]~1.4–3 across various sources (Lefloch et al., 2016; Rivilla et al., 2018; Bernal et al., 2021).

Besides the discovery of new chemical species in the ISM through spectroscopic techniques, astrochemistry also relies on laboratory experimental work and computational models. Laboratory experiments replicate interstellar physical conditions to obtain rotational spectra for known molecular compositions. Computational models, on the other hand, integrate data from astronomical measurements and laboratory experiments, providing a framework to hypothesize the underlying physical-chemical processes responsible for the observed molecular abundances. In this context, astrochemical models such as UCLCHEM (Holdship et al., 2017) or Nautilus (Ruaud et al., 2016) have been developed to simulate the time evolution of abundances for numerous interstellar species, which interact through extensive networks of chemical reactions under various physical conditions and energetic physical processes (e.g., UV/cosmic rays irradiation, protostellar heating or stellar feedback in the form of high-velocity

winds inducing shock waves) in both the gas phase and on dust grain surfaces. These models typically consist of applying the law of mass action to all reactions in the chemical network and solving numerically the associated systems of ordinary differential equations. However, the precision of mass-action kinetics depends heavily on the accuracy of the reaction rate coefficients, which remain poorly determined for the majority of reactions.

In the past decade, it has become clear that astrochemical models present severe limitations in reproducing the [PO]/[PN] ratios observed across the ISM (predicted ratios < 1 versus the observed ratios of ~1.4–3; see e.g., Aota and Aikawa (2012), Jiménez-Serra et al. (2018), Chantzios et al. (2020), Sil et al. (2021)). To get insight into this problem, in a previous work, we conducted a detailed analysis of the chemistry of phosphorus in molecular clouds (Fernández-Ruz et al., 2023). This analysis employed a theoretical approach integrating complex networks, sensitivity analysis, and Bayesian statistics to model 14 key chemical reactions. Unlike traditional numerical approaches in astrochemical modeling, our theoretical method provided deep insights into the role of each reaction in the formation of PO and PN, resolving the discrepancies between modeled and observed abundance ratios. Additionally, this study revealed a strong hierarchical structure in the kinetic parameters of the system, where the formation of PO and PN in molecular clouds was insensitive to a large number of parameter combinations but highly dependent on a selected few. This intrinsic parameter insensitivity aligns very well with the concept of a “sloppy model”, where system behavior is governed by a few “stiff” parameter combinations, while others, referred to as “sloppy directions”, exhibit extreme insensitivity to large-scale fluctuations (Brown and Sethna, 2003; Quinn et al., 2023).

Sloppy models are prime candidates for model reduction techniques. In general, these techniques aim to simplify large-scale dynamical systems through the elimination of unnecessary parameters while preserving essential characteristics (Antoulas, 2005). The primary techniques can be categorized into (i) eigenvector projection methods (using Singular Value Decomposition) and (ii) moment matching methods. Some eigenvector projection methods *diagonalize* the system and truncate those states that are difficult to control and observe (Moore, 1981), while others aim to minimize an error measure of the system (Glover, 1984) or separate fast and slow timescales (Gugercin and Antoulas, 2004). On the contrary, moment-based methods focus on reducing the complexity while approximating accurately some statistics (often central moments estimated from the data). Overall, both categories of methods strive to achieve a balance between computational efficiency and accuracy, making them suitable for large-scale models but relying too strongly on linearization or on global measurements. Alternatively, M.K. Transtrum and collaborators, following a different approach, developed a decade ago the Manifold Boundary Approximation Method (MBAM) (Transtrum et al., 2011; Transtrum and Qiu, 2014; Transtrum and Qiu, 2016; Quinn et al., 2023). This elegant algorithm is based on information geometrical arguments, balancing the data with the complexity of the model exploiting the low sensitivity to some combinations of parameters.

In this paper, we introduce a methodology for reducing and simplifying models while retaining their predictive accuracy inspired by the MBAM reduction technique and apply it specifically

to the phosphorus astrochemistry network. The proposed method, the Fisher Information Spectral Reduction method (hereafter, FISR method), is conceptually simpler than the MBAM and thus can be implemented by an easier and more efficient algorithm. It performs iterative dimensional reduction steps, progressively decreasing the number of model parameters. The final outcome is a reduced model with significantly fewer parameters, in which unnecessary complexity is removed.

Applying the FISR algorithm to the 14-reaction model of phosphorus chemistry in molecular clouds presented in Fernández-Ruz et al. (2023), we demonstrate that the observed abundances of PO and PN can be explained by a much simpler model comprising just 3 reactions and 3 parameters. This simplified model not only identifies the key reactions governing PO and PN formation but also offers a deeper comprehension of the hierarchical structure of the parameter space. Therefore, the goal of this work is not to obtain a model that improves the predictions of existing astrochemical models, but rather to derive simpler and interpretable models that maintain comparable predictive accuracy while eliminating unnecessary complexities. Our findings highlight that the complexity of a model should correspond to the complexity of the available knowledge with which it is constructed. This principle—also known in the sciences as *Occam's razor*—, is broadly applicable beyond astrochemistry to other disciplines reliant on modeling.

The organization of this paper is as follows. In Section 2 we present the mathematical foundation of the complexity reduction method introduced here, the FISR method, and outline the algorithmic steps for its general implementation. In Section 3 we apply the FISR method to the phosphorus chemistry in the interstellar medium, uncovering the underlying chemical network dynamics that governs the formation of PO and PN across short and long timescales. In Section 4, we discuss the implications of the main findings for future phosphorus astrochemical research and examine the strengths and limitations of the FISR method within the context of complexity reduction methods.

2 Methods: the FISR algorithm

The method presented in this work is inspired by the Manifold Boundary Approximation Method (MBAM), originally developed and published by Transtrum and Qiu (2014), and shares its basic principles and primary objective. The reasoning behind this method is that models often do not respond significantly to changes in certain parameters. By applying concepts from information geometry, this approach identifies specific trajectories across the statistical manifold that represent meaningful combinations of the parameters. These trajectories follow the so-called sloppy directions, which are directions in the parameter space that are not informative and along which large parameter changes produce only a minimal impact on the model output. The term *boundary* in the name of the method connects with the practical observation made by the authors that, except in rare pathological cases, the topology of an N -dimensional manifold contains boundaries of $N - 1$ dimensions. This is the reason why one can drop one combination of parameters safely when a trajectory over the manifold reaches one boundary,

while preserving the model's predictive power (Transtrum and Qiu, 2014).

Our proposed algorithm, the Fisher Information Spectral Reduction (FISR) method, successfully achieves the same goals as MBAM but differs significantly from it in its technical details and algorithmic procedures. We name it FISR method because it is based on the eigenspectrum of the Fisher Information Matrix (FIM), which quantifies the amount of information a dataset carries about the model parameters, but also provides the directions on the statistical manifold carrying less information (the sloppiest directions). Since high information content is associated with strong parameter influence, the eigenspectrum of the FIM encodes the sensitivity of model outputs to changes in parameter values. In this Section, we present the mathematical framework of the FISR method and outline its algorithmic implementation from a general perspective, making it applicable to any sloppy model.

2.1 Theoretical framework and definitions

Let us consider a model $\mathbf{y} = \mathbf{y}(\boldsymbol{\theta})$ with a set of N parameters $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_\mu, \dots, \theta_N\}$ (see Table 1 for a summary of all the symbols used in this Section). The M outputs of the model are $\mathbf{y} = \{y_1, \dots, y_m, \dots, y_M\}$, where m typically stands for time values of the evolution of one or more quantities of interest. The *optimal set* of outputs \mathbf{y}^{opt} is defined as the most reliable output values and can either be (i) real data obtained from measurements or observations or (ii) synthetic data, that is, predicted values generated by a model that is considered sufficiently accurate by the user. For each set of parameters $\boldsymbol{\theta}$ and one output (or prediction point) $y_m(\boldsymbol{\theta})$, we define the residual $r_m(\boldsymbol{\theta})$ as the difference between the *optimal* value y_m^{opt} and the output $y_m(\boldsymbol{\theta})$, that is,

$$r_m(\boldsymbol{\theta}) = y_m^{opt} - y_m(\boldsymbol{\theta}). \quad (1)$$

In addition, we define the cost function as

$$C(\boldsymbol{\theta}) = \sum_{m=1}^M (y_m^{opt} - y_m(\boldsymbol{\theta}))^2. \quad (2)$$

The cost quantifies how far the outputs obtained with parameter set $\boldsymbol{\theta}$, i.e., $\mathbf{y}(\boldsymbol{\theta})$, are from the optimal \mathbf{y}^{opt} and is therefore used to assess the reliability of a parameter set $\boldsymbol{\theta}$. Furthermore, the sensitivity of the model residuals to changes in the parameters is arranged in the Jacobian matrix \mathbf{J} , with elements

$$J_{m\mu} = \frac{\partial r_m}{\partial \theta_\mu}. \quad (3)$$

Finally, the Fisher Information Matrix (FIM) \mathbf{g} has a key role in our method, and its elements are defined as

$$g_{\mu\nu} = \sum_{m=1}^M \frac{\partial r_m}{\partial \theta_\mu} \frac{\partial r_m}{\partial \theta_\nu}. \quad (4)$$

In matrix form, the FIM and the Jacobian are related through $\mathbf{g} = \mathbf{J}^T \mathbf{J}$, being \mathbf{J} (Equation 3) a matrix of size $M \times N$ and \mathbf{g} (Equation 4) a matrix of size $N \times N$. As the FIM \mathbf{g} is square, symmetrical and positive semi-definite, all its eigenvalues are real and non-negative, and quantify how strongly perturbations along the corresponding

TABLE 1 Notation used in the FISR method description.

Description	Notation
Model parameters/Parameter space	θ
Element of the parameters vector	θ_μ
Prediction	$y(\theta)$
Prediction point	$y_m(\theta)$
Model	$y = y(\theta)$
Residual of a prediction point for a set of parameters	$r_m(\theta)$
Cost for a set of parameters	$C(\theta)$
Time point in the path	τ_i
Path	$\theta(\tau)$
Parameter values at one point in the path	$\theta(\tau_i)$
FIM evaluated at one point in the path	$\mathbf{g}(\theta(\tau_i))$
Velocity vector at one point in the path	$\mathbf{v}(\tau_i)$

eigenvector directions in the parameter space affect the model output y . Note that, as the Jacobian represents the change in the output with respect to parameter perturbations, it is a matrix of sensitivities (Saltelli, 2008). Thus, another interpretation of the FIM is the matrix of second-order sensitivities (also known as parameter synergies) (Faro et al., 2019).

In cases where the parameters are positive by definition and/or their scales are significantly different, it will be very helpful to calculate \mathbf{J} and \mathbf{g} after a re-parametrization of the model such that $\tilde{\theta}_\mu = \ln(\theta_\mu)$, as was already stated in the original MBAM algorithm (see, for example, Transtrum and Qiu (2016)). This re-parametrization is aimed at weighing the order of magnitude of the parameter rather than its value, so the method does not necessarily bias the elimination procedure toward smaller values.

In practice, sloppy models are those whose FIM eigenvalues span along many orders of magnitude, from eigenvalues close to zero, whose associated eigenvectors pinpoint the parameter directions that hardly affect the system—the sloppy directions—, to large eigenvalues, whose eigenvectors denote directions of high parameter sensitivity. We will exploit this feature of the FIM for parameter reduction in the next Subsection.

2.2 Parameter reduction algorithm

The FISR algorithm is implemented through successive dimensionality reduction steps, reducing the original model iteratively until no further simplifications are possible. Each reduction step takes an initial model as the input and produces a simplified version as the output. This iterative process is carried out by an algorithm structured in the following three stages:

Stage 1. Navigation through the parameter space

Considering an initial model $y(\theta)$, each set of parameters θ represents one point in the parameter space associated with $y(\theta)$. This set has an associated FIM $\mathbf{g}(\theta)$ and cost $C(\theta)$. The parameter navigation consists of a path parameterized by the variable τ , $\theta(\tau)$, that starts from the initial point $\theta(\tau_0)$ (where the cost is minimum), and evolves according to

$$\theta(\tau_{i+1}) = \theta(\tau_i) + \mathbf{v}(\tau_i) \delta\tau, \quad (5)$$

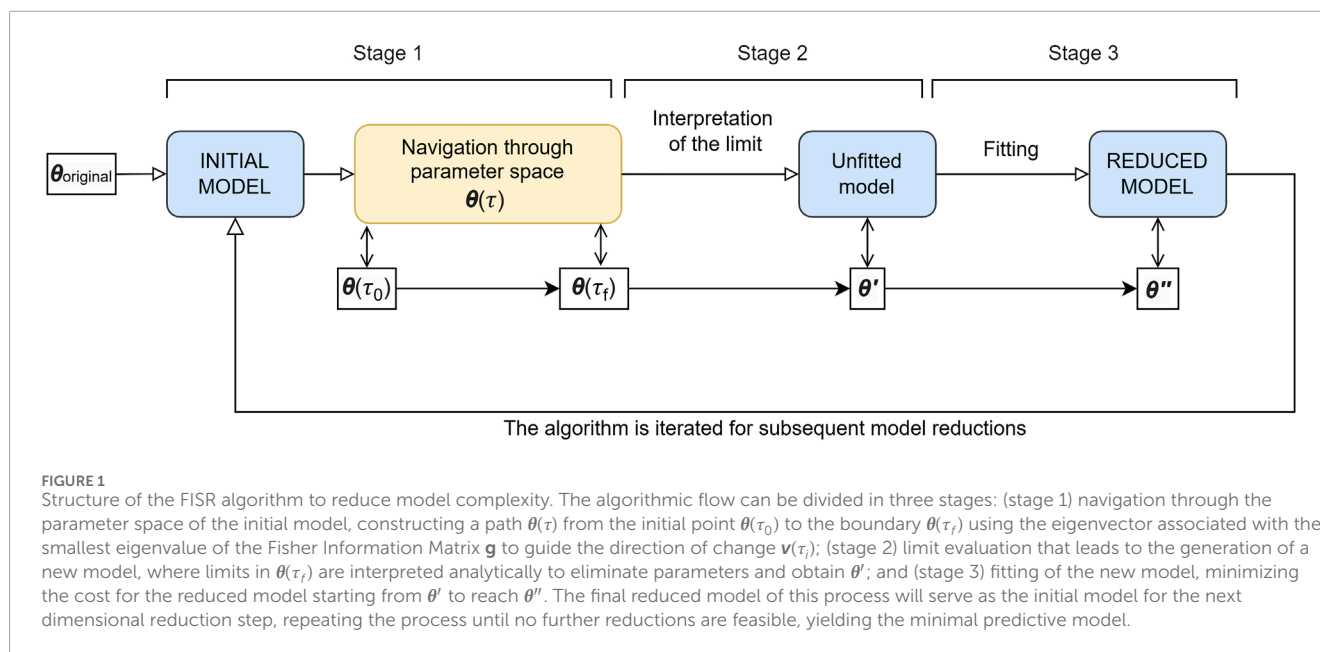
where $\mathbf{v}(\tau_i)$ is the velocity vector at iteration i and $\mathbf{v}(\tau_i) \delta\tau$ denotes a small size step. The velocity $\mathbf{v}(\tau_i)$ is a L2-normalized vector in the direction of the eigenvector corresponding to the smallest eigenvalue of the FIM $\mathbf{g}(\theta(\tau_i))$ (namely, the sloppiest direction). Since the orientation of the eigenvectors is undefined, among the two possible ones, in each step τ_i the orientation of $\mathbf{v}(\tau_i)$ is chosen such that it verifies that the angle of the velocities along the geodesic between consecutive steps is lower than $\pi/2$, i.e., that $\mathbf{v}(\tau_i) \cdot \mathbf{v}(\tau_{i-1}) > 0$, ensuring that the path $\theta(\tau)$ does not go backward locally and therefore allowing a fast displacement from the original point¹. This definition of $\mathbf{v}(\tau_i)$ based on the FIM certifies that the small step $\mathbf{v}(\tau_i) \delta\tau$ has the minimum possible effect on the model output and thus on the cost. In principle, the smoothness of the model and the lack of global conserved quantities guarantee that the path $\theta(\tau)$ reaches a boundary (Transtrum and Qiu, 2014). For instance, if a parameter is a positive real number, the theoretical limits would be zero and infinity (Transtrum et al., 2011), so the manifold boundary would correspond to a region where one or more parameters take those limiting values. In practice, the path is considered finished when the tendency of at least one parameter toward a limit is persistent over time. For consistency, we denote $\theta(\tau_f)$ as the set of parameters after the last iteration of Equation 5.

Stage 2. Limit evaluation and generation of a new model

As mentioned above, stage 1 ends when one or more θ_μ reach a boundary (limiting value). In stage 2, a new model is constructed by considering the physical meaning of the limiting values of these parameters. For the typical case where parameters are positive real numbers and only one or two parameters reach a boundary, the possible scenarios are:

- A parameter approaches zero in stage 1. This may indicate that the microscopic process it represents has become negligible in the system with the new values of the parameters, for example, because it is much slower than the rest of the processes. In such cases, the new model excludes this process and the corresponding parameter.
- A parameter approaches infinity in stage 1. This is typical when the parameters are reaction rate coefficients, and implies that

¹ Note that with this criterion, the orientation of $\mathbf{v}(\tau_0)$ remains undefined, meaning that two valid paths coexist and one can choose any of them arbitrarily.



a mechanism is almost *instantaneous* in comparison to others. This provides an *adiabatic elimination* of the fast timescale, and this mechanism can now be described by a single process with the proper change in the initial conditions.

- (iii) Two or more parameters tend to infinity or one to infinity while the other to zero in stage 1. Here, a new parameter that combines them but remains finite is introduced in the model.

Note, however, that more complicated scenarios are possible, and more than two parameters may tend to zero or infinity in the same reduction step. In any case, the interpretation of the limit to generate the new model is a task that cannot be automated and must be done by hand because it requires knowledge about the processes that are modeled and the physical meaning of the parameters. Note that the new model is now evaluated in the parameter set θ' , which is obtained by implementing the changes in $\theta(\tau_f)$ and therefore has fewer parameters than the initial model.

Stage 3. Fitting of the new model

Once the new model is defined, it is fitted by finding the parameter set θ'' that minimizes the cost. The initial guess is θ' . The outcome of stage 3 is a reduced model with fewer parameters than the initial one, but due to the methodology presented in stages 1, 2, and 3, the new reduced model achieves predictions as close as possible to those of the initial model.

Consecutive dimensional reduction steps. The whole process of parameter reduction starts with the original model y_0 and is iterated several times, where each loop is defined as a dimensional reduction step s . The complete algorithm structure is shown in Figure 1. At each step s , model y_{s-1} is reduced, yielding model y_s ². A reduction

step can be performed as long as there is a sloppy direction. We know that no further reductions are possible when the cost becomes so large that the output of the simplified model is too different from that of the original model. The minimal model will be the one that retains its predictive power with the smallest possible number of parameters.

3 Results

In this Section, we apply the FISIR method introduced in this work to a phosphorus astrochemistry model comprising 14 chemical reactions involving 7 P-bearing molecules proposed by Fernández-Ruz et al. (2023). The parameters subject to reduction are the kinetic parameters, specifically the rate coefficients k_μ of the 14 reactions. Our focus is on the abundances of PO and PN, the most abundant P-bearing molecules detected so far in star-forming regions (Lefloch et al., 2016; Rivilla et al., 2020; Fontani, 2024). We perform 9 dimensional reduction steps, 8 of which preserve with high precision the model's ability to predict the evolution of PO and PN abundances within the time range from 10^4 to 10^5 yrs. These timescales are typical of molecular outflows in star-forming regions where PO and PN have been detected (e.g., Lefloch et al. (2016), Rivilla et al. (2020), Bernal et al. (2021), Fontani (2024)), and are the ones usually explored with astrochemical models (see e.g., Jiménez-Serra et al. (2018)). The FISIR method results in a final minimal model with only 3 rate coefficients.

3.1 Original model for the chemical evolution of phosphorus in the ISM

We adopt the phosphorus chemistry model for the ISM from our recent publication (Fernández-Ruz et al., 2023) as the original model of the complexity reduction process. We opt for this model

² From now on, we will refer to model y_s as model s for simplicity.

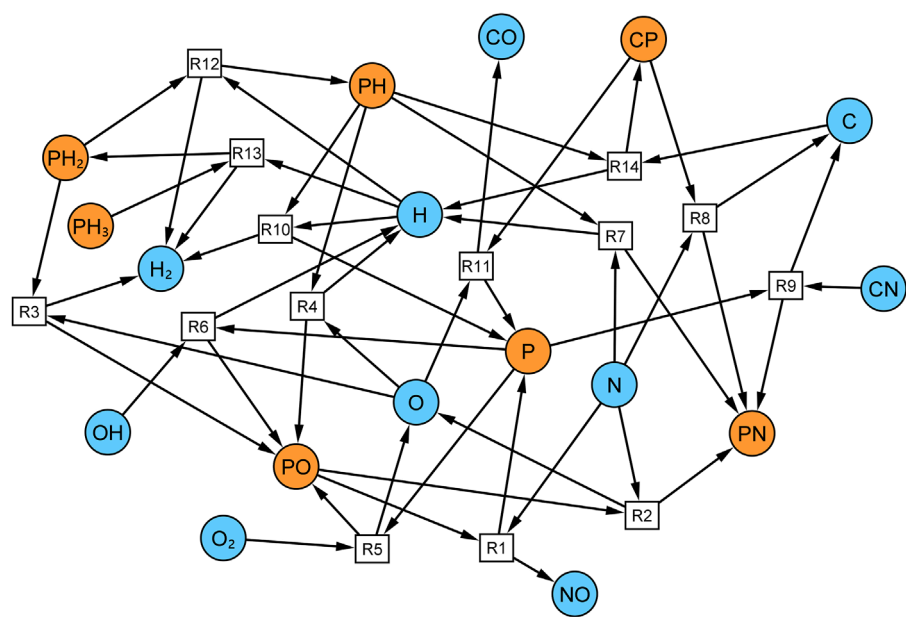


FIGURE 2
Bipartite chemical network representing the 14 chemical reactions involved in the phosphorus chemistry in the interstellar medium studied in this work. Orange circular nodes correspond to P-bearing species, whose abundance evolves over time as described by Equation 6. The rest of the species are represented as blue circular nodes. Square nodes (labeled R1–R14) denote chemical reactions, with directed links pointing from reactants to reaction nodes, and from reaction nodes to products.

TABLE 2 Reactions and their rate coefficients. (a) The rate coefficients of these reactions were adjusted with Bayesian inference in Fernández-Ruz et al. (2023). (b) The Arrhenius parameters of these rate coefficients come from theoretical quantum chemical calculations.

Reaction n°	Reaction	k_{μ} (cm ³ s ⁻¹) (T = 100 K)	Source
1	N + PO → P + NO	1.81 × 10 ⁻¹²	Wakelam et al. (2012) (a)
2	N + PO → PN + O	2.27 × 10 ⁻¹²	Wakelam et al. (2012) (a)
3	O + PH ₂ → PO + H ₂	4.00 × 10 ⁻¹¹	Wakelam et al. (2012)
4	O + PH → PO + H	1.00 × 10 ⁻¹⁰	Wakelam et al. (2012)
5	P + O ₂ → PO + O	4.38 × 10 ⁻¹⁶	de la Concepción et al. (2024) (b)
6	P + OH → PO + H	1.91 × 10 ⁻¹⁰	García de la Concepción et al. (2021) (b)
7	N + PH → PN + H	1.06 × 10 ⁻¹⁰	Gomes et al. (2023) (b)
8	N + CP → PN + C	5.55 × 10 ⁻¹¹	Wakelam et al. (2012)
9	P + CN → PN + C	5.55 × 10 ⁻¹¹	Wakelam et al. (2012)
10	H + PH → P + H ₂	3.29 × 10 ⁻¹²	Charnley and Millar (1994) (a)
11	O + CP → P + CO	4.00 × 10 ⁻¹¹	Wakelam et al. (2012)
12	H + PH ₂ → PH + H ₂	2.58 × 10 ⁻¹²	Charnley and Millar (1994)
13	H + PH ₃ → PH ₂ + H ₂	2.89 × 10 ⁻¹⁴	Charnley and Millar (1994)
14	C + PH → CP + H	7.50 × 10 ⁻¹¹	Wakelam et al. (2012)

TABLE 3 Initial gas-phase abundances that apply in the original model, extracted from [Fernández-Ruz et al. \(2023\)](#). Some initial abundances are missing because they are not needed as they do not appear in the models' equations. (a) In cases where the source provided two values or we considered two sources, we used the geometric mean. (b) Up to date, CP has not been detected in the ISM, but it has been detected in a circumstellar shell envelope by [Guelin et al. \(1990\)](#). Thus, in our model we consider that CP is present but we fix its initial value to 10^{-15} so it is sufficiently below the detection limit ($\sim 10^{-12}$). (c) The value is manually set to one (for H) and zero (for PO and PN) in agreement with the models' configuration. (d) As proposed in other works ([Jiménez-Serra et al., 2018](#); [Fernández-Ruz et al., 2023](#)), we assume that P (cosmic abundance of 2.57×10^{-7}) is depleted by a factor of 100 and that 50% is hydrogenated on the surface of dust grains forming PH, PH₂ and PH₃. The rest remains as atomic P. For simplicity, PH, PH₂ and PH₃ are formed in equal amounts, leading to the initial abundances shown in the table.

Species	Initial abundance (n/n_H)	Reference
C	2.69×10^{-4}	Jiménez-Serra et al. (2018)
CN	5.92×10^{-10}	Agúndez and Wakelam (2013) (a)
CO	-	N/A
CP	1.00×10^{-13}	Guelin et al. (1990) (b)
H	1	N/A (c)
H ₂	-	N/A
N	6.76×10^{-5}	Jiménez-Serra et al. (2018)
NO	-	N/A
O	4.90×10^{-4}	Jiménez-Serra et al. (2018)
O ₂	6.04×10^{-7}	Larsson et al. (2007) , Goldsmith et al. (2011)
OH	1.00×10^{-8}	Rugel et al. (2018) (a)
P	1.29×10^{-9}	Jiménez-Serra et al. (2018) (d)
PH	4.28×10^{-10}	Jiménez-Serra et al. (2018) (d)
PH ₂	4.28×10^{-10}	Jiménez-Serra et al. (2018) (d)
PH ₃	4.28×10^{-10}	Jiménez-Serra et al. (2018) (d)
PN	0	N/A (c)
PO	0	N/A (c)

because it has been proven to accurately describe the [PO]/[PN] ratios of the whole phosphorus network predicted by UCLCHEM, while remaining sufficiently small to apply the FISR method. The system, whose associated chemical network is shown in [Figure 2](#), describes the evolution of phosphorus chemistry in a star-forming region over 10^5 yrs. For the simulations, we assume constant physical conditions with a temperature of $T = 100$ K and a cloud H density of $n_H = 10^4 \text{ cm}^{-3}$, typical of regions affected by shocks associated with molecular outflows and where PO and PN have been detected (see e.g., the works of [Lefloch et al. \(2016\)](#) or [Rivilla et al. \(2020\)](#)). The model represents the dynamics resulting from the application of the law of mass action to a network of 14 chemical reactions with their corresponding 14 reaction rate coefficients k_μ , listed in [Table 2](#), and involving 17 chemical species with initial conditions provided in [Table 3](#). Among these species, there are 7 P-bearing molecules: P, PH, PH₂, PH₃, CP, PO, and PN. Importantly, most of the reaction rate coefficients k_μ were calculated using the Arrhenius parameters extracted from the KIDA database ([Wakelam et al., 2012](#)). However, three of these coefficients were

later adjusted for $T = 100$ K using Bayesian inference to obtain PO and PN predictions that match the observed values (see reactions 1, 2 and 10 in [Table 2](#) and [Fernández-Ruz et al., 2023](#)). More precisely, the model predicts $[PO]/[PN] = 3.3$ at $t = 10^4$ yrs, in agreement with [Lefloch et al. \(2016\)](#), [Rivilla et al. \(2018\)](#), [Bernal et al. \(2021\)](#), which reported [PO]/[PN] ratios equal to 2.8, 1.4 and 2.6, respectively. Finally, we assume that P is initially adsorbed onto dust grains, from which 1% is released into the gas phase ([Jiménez-Serra et al., 2018](#); [Fernández-Ruz et al., 2023](#)). However, note that the actual depletion factor of P in the ISM is still a debated question ([Fontani, 2024](#)). In addition, we set a P-hydrogenation fraction f_p of 0.5, meaning that 50% of the released P is hydrogenated, and the rest remains as atomic P. For simplicity, PH, PH₂ and PH₃ are formed in equal amounts, leading to the initial abundances shown in [Table 3](#), but let us remark that there are currently no observational information about the abundance of P, PH, PH₂ and PH₃ in star-forming regions, and therefore, we can only provide reasonable guesses for these parameters. Interestingly, though, we have recently shown that the evolution of the system is in practice independent

of the P-hydrogenation fraction f_p for $T = 100$ K beyond 10^4 yrs, that is, in the time range of interest for this paper (Fernández-Ruz et al., 2023).

We refer to the system of ordinary differential equations (ODEs) with 17 variables, derived from applying the law of mass action to the astrochemical model plotted in Figure 2, as the *total system*. This system is mathematically intractable and can only be solved using numerical methods. However, the *linear effective system* is a linearized approximation of the full system that can be theoretically solved under certain approximations for the 7 P-bearing species. The linearization is possible when the non P-bearing species are treated as constants, an assumption justified by their significantly higher abundances, as shown in Table 3 (see Supplementary Section S1.1 for details). The linear effective system solution consists of explicit equations for the time-evolving abundances:

$$[X_p] = \sum_{q=1}^{p-1} \left(\frac{C_{pq}}{r_p - r_q} e^{-r_q t} \right) + C_{pp} e^{-r_p t}, \quad (6)$$

where $[X_p]$, with $p = 1$ to 7, represents the abundance of the P-bearing species PH_3 , PH_2 , PH , CP , P , PO , and PN , respectively. The constants C and r depend on the reaction rate coefficients and initial abundances, with their explicit expressions provided in Supplementary Section S1.2. Also, note that, for $p = 1$, the summation corresponds to an empty sum; thus, its value is zero. This theoretical solution of the linear effective system approximates the numerical solution of the total system for the P-bearing species with average relative errors of $\sim 1\%$ for $T = 100$ K, while being the former more than 5 orders of magnitude computationally faster than the latter (Fernández-Ruz et al., 2023).

The sensitivity analysis presented in Fernández-Ruz et al. (2023) revealed that the abundances of PO and PN at times $t = 10^3$, $t = 10^4$, and $t = 10^5$ yrs and for $T = 100$ K exhibit negligible dependence on some of the rate coefficients k_μ and large sensitivity on a few of them. This indicates that the model shows strong parameter hierarchy or sloppiness, making it particularly well-suited for parameter reduction using the FISR method, with the rate coefficients k_μ as the 14 parameters and PO and PN as the outputs. We do not consider the remaining P-bearing molecules as outputs because they have not been detected in star-forming regions (Fontani, 2024). Based on this, and following the notation defined in Section 2.1, we call original model—that will act as a model 0—to $\mathbf{y} = \mathbf{y}(\boldsymbol{\theta} = \mathbf{k})$, where $y_m = y_m(k_1, \dots, k_{14})$ and $m = 1, \dots, 50$ enumerates the abundances of PO and PN given by Equation 6, evaluated at 25 time points uniformly distributed between $t = 10^4$, and $t = 10^5$ yrs—the time range typically observed in real data (see above). Finally, we need to obtain the optimal prediction \mathbf{y}^{opt} used in the calculation of the residuals and the cost (Equations 1, 2). Since we lack a temporal series of observational data points for the abundances of PO and PN in a molecular cloud—real data correspond to a single snapshot of the cloud's time evolution—we generate synthetic data from the original model, considering that $\mathbf{y}^{opt} = \mathbf{y}(\mathbf{k}_{\text{original}})$, where $\mathbf{k}_{\text{original}}$ is the set of rate coefficients values showed in Table 2. By construction, the cost of the original model, with $\mathbf{k}_{\text{original}}$ as input, is zero.

3.2 Toward the simplest model compatible with the chemical evolution of phosphorus in the ISM

3.2.1 Dimensionality reduction steps

Here, we apply the FISR algorithm to the original model of the phosphorus chemistry in the ISM through a series of iterative dimensionality reduction steps. As explained in Section 2.2, each reduction step s in the algorithm consists of (i) the parameter space navigation in the model $s - 1$ until a boundary is reached, (ii) the evaluation of the limit and model reinterpretation to obtain model s , with less parameters, and (iii) the fitting of model s . Therefore, a new model with fewer parameters and, in some cases, fewer variables is derived at each reduction step s . We adopt the re-parametrization $\tilde{k}_\mu = \ln(k_\mu)$ justified in Section 2.1, and also the abundances of PO and PN are log-transformed in the calculation of the cost and the FIM. The fitting of models is done by finding the set of rate coefficients \mathbf{k} that minimizes the cost with the Nelder-Mead method (Lagarias et al., 1998). In summary, our goal now is to evaluate the performance of the new models and find the simplest set of reactions that can yield time-evolution abundance curves of PO and PN from 10^4 to 10^5 yrs that are indistinguishable from those obtained from the original model. We remind the reader that we focus on timescales between 10^4 to 10^5 yrs because these are the relevant timescales for the star-forming regions where PO and PN have been detected (Lefloch et al., 2016; Rivilla et al., 2020; Fontani, 2024). Figure 3 shows the evolution of the set of rate coefficients \mathbf{k} along the parameter space at each reduction step, obtained in stage 1 of the FISR algorithm. Each path has been computed for 10^5 iterations, far enough to appreciate a clear tendency in which at least one rate coefficient k_μ goes to zero or infinity. The cases where a parameter k_μ goes to zero are easy to interpret: the corresponding reaction can be removed from the model since the FISR algorithm proves that it has a negligible effect on the cost, which ultimately means that the reaction is not relevant for the final abundances of PO and PN. Therefore, the new model s has one less rate coefficient and is obtained by removing that reaction from model $s - 1$. On the contrary, the cases where a k_μ goes to infinity must be interpreted individually.

After each reduction step, a new model is built, and this process is repeated along 9 dimensionality reduction steps. Figure 4 shows the time evolution of the abundances for the P-bearing species of every new model s compared to the original model. For completeness, Supplementary Section S2 shows the same comparison for the ratio $[\text{PO}]/[\text{PN}]$, a relevant quantity in phosphorus astrochemistry, and Supplementary Section S3 compiles the initial conditions and chemical reactions that constitute every model, from the 14 reactions of the original model to the 2 reactions of model 9, the most simplified one.

As we can see in Figures 3A–C, corresponding to reduction steps 1, 2 and 3, respectively, the reaction rate coefficients k_5 , k_7 and k_{11} tend to zero. In consequence, reactions 5 ($\text{P} + \text{O}_2 \rightarrow \text{PO} + \text{O}$), 7 ($\text{N} + \text{PH} \rightarrow \text{PN} + \text{H}$) and 11 ($\text{O} + \text{CP} \rightarrow \text{P} + \text{CO}$) can be interpreted as infinitely slow in comparison to the rest, and can be sequentially removed, yielding models 1, 2 and 3. For reaction 5, this is consistent with new quantum chemical calculations of this rate constant, which show that its value at 100 K is very small (de la Concepción et al., 2024). In Figures 4A–C we see that PN and CP curves differ between the original

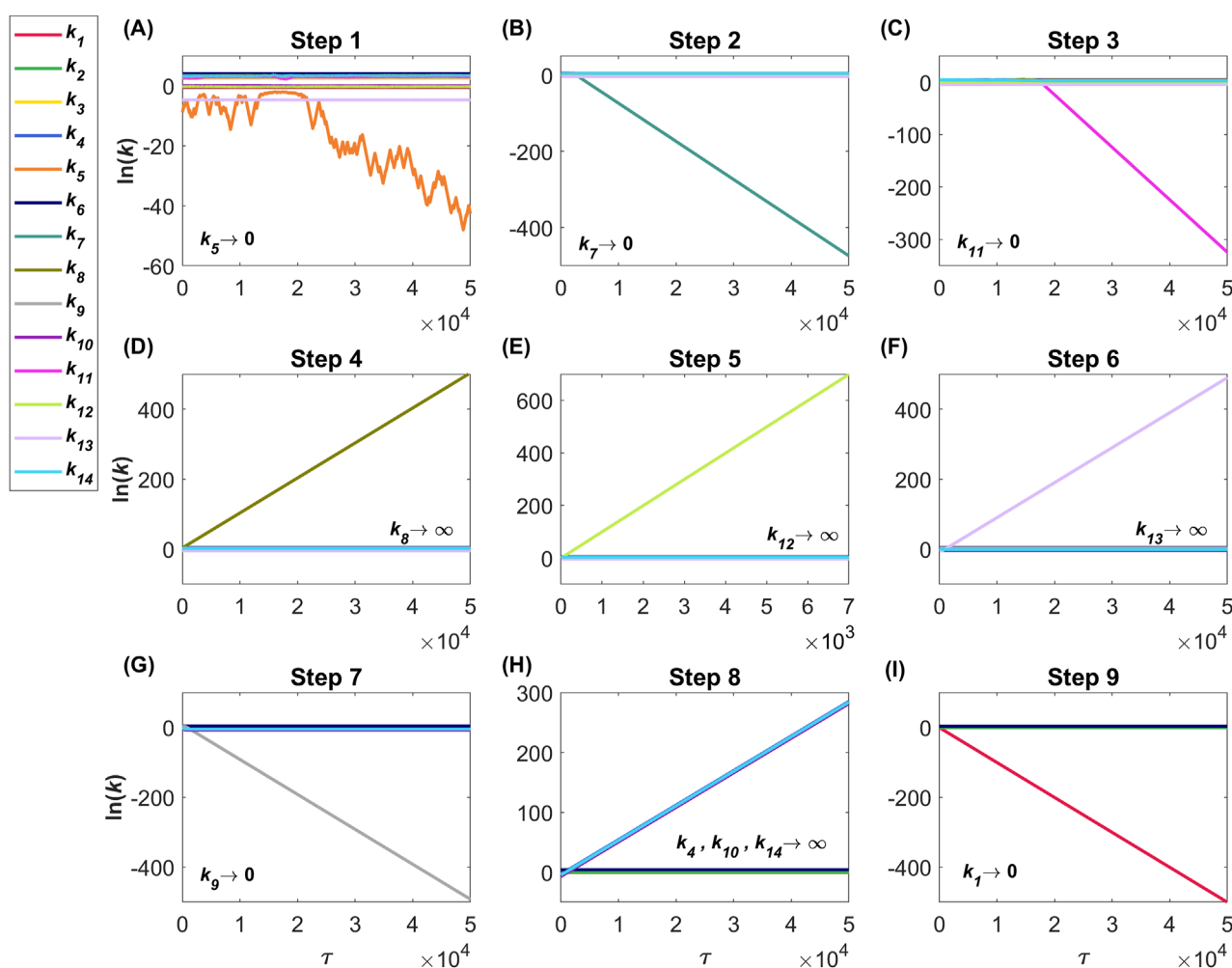


FIGURE 3

Evolution with τ of the rate coefficients k_μ of the phosphorus astrochemical system in the ISM along the path $\theta(\tau) = \mathbf{k}(\tau)$ obtained in stage 1 of the FISIR algorithm, for 9 dimensionality reduction steps. τ is the time step in the navigation through the parameter space. The path starts at the initial point $\mathbf{k}(0)$ where the cost is minimum and tends to a boundary where one or more k_μ reach zero or infinity. In steps 1, 2, 3, 7 and 9 (panels A, B, C, G, I), the rate coefficients k_5 , k_7 , k_{11} , k_9 and k_1 , respectively, approach zero. In steps 4, 5 and 6 (panels D, E, F), the rate coefficients k_8 , k_{12} and k_{13} tend to infinity. In step 8 (panel H), the rate coefficients k_4 , k_{10} and k_{14} tend to infinity in parallel. The number of iterations is 5×10^4 for all reduction steps except for step 5, which only reaches $\sim 7,000$ iterations due to computational limitations arising from exceedingly large values.

model (O.M., dashed lines) and the new models (solid lines) for low times, but the PO and PN time-evolution abundance profiles after $t = 10^3$ yrs remain indistinguishable. Noticeably, reduction step 1 shows a strongly noisy behavior for k_5 because the equations of the original model present a structural non-identifiability regarding parameters k_5 and k_6 , as the mathematical solutions of the system depend on these two parameters exclusively through the sum $k_5[\text{O}_2]_0 + k_6[\text{OH}]_0$. Structural non-identifiability arises when the model's equations contain redundant parameters or symmetries, making it impossible to uniquely determine two or more parameters from the data, regardless of its quality.

Figure 3D shows that the rate coefficient of reaction 8 ($\text{N} + \text{CP} \rightarrow \text{PN} + \text{C}$) approaches infinity in the reduction step 4. This means that we can define a new model—model 4, see [Supplementary Section S3](#)—where this reaction is infinitely fast in comparison to the rest, so the initial abundance of CP is added to the initial abundance of PN,

removing CP as a variable of the system³. Besides reaction 8, only reaction 14 ($\text{C} + \text{PH} \rightarrow \text{CP} + \text{H}$) involves CP in model 3, so it can be merged with reaction 8, yielding the new reaction 14 ($\text{C} + \text{PH} + \text{N} \rightarrow \text{PN} + \text{H} + \text{C}$) that accounts for the 2-step process: $\text{C} + \text{PH} \rightarrow \text{CP} + \text{H}$ followed by $\text{N} + \text{CP} \rightarrow \text{PN} + \text{C}$. Since in this case the first reaction is the rate-determining step, the rate law of the new reaction 14 is $v = k_{14}[\text{C}][\text{PH}]$. Figure 4D shows that, despite model 4 lacking CP (green solid line is absent), new reaction 14 can capture the growth in the abundance of PN from the original model.

³ Recall that the non P-bearing species are assumed to be constant due to their higher abundances and the P-bearing species become the only dynamical variables. Therefore, for all the cases in this Subsection, we only consider changes in the initial conditions of the P-bearing species, as the changes in the non P-bearing species are negligible.

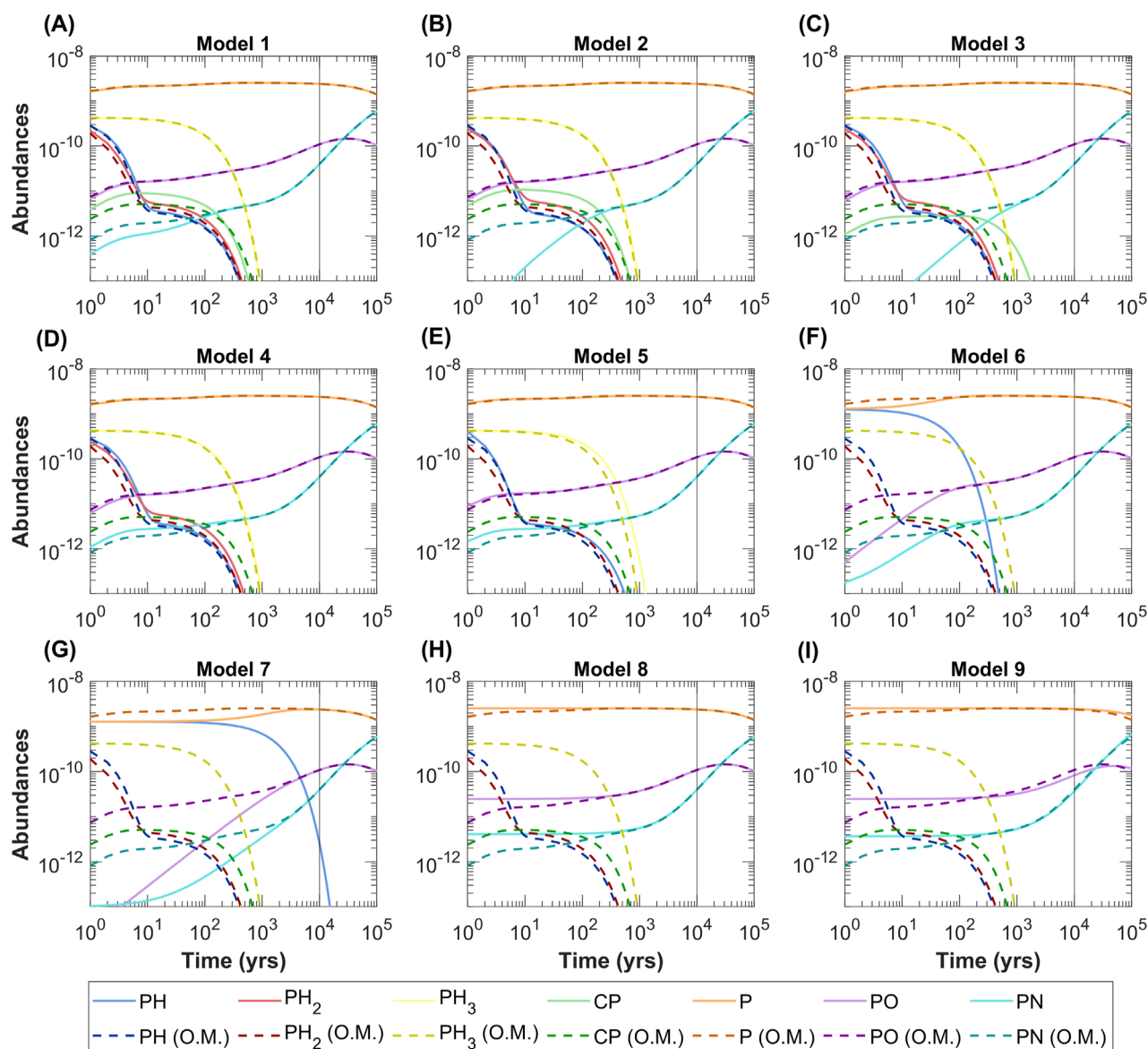


FIGURE 4

Evolution with time of the abundances relative to H of the P-bearing molecules for models 1 to 9 (solid lines), obtained in stage 3 of the FISIR algorithm, and for the original model (O.M., dashed lines; Fernández-Ruz et al., 2023) of the phosphorus astrochemistry in the ISM. Models 1–9 (panels A–I) are progressively simpler approximations of the original model. Although only abundances of PO and PN beyond 10⁴ yrs (vertical black lines) were used as input in the reduction algorithm, models 1–8 reproduce precisely the abundances of PO and PN of the original model beyond a few 100 years. On the contrary, the disagreement between model 9 and the original model can be seen with the naked eye.

In reduction step 5, the rate coefficient of reaction 12 ($\text{H} + \text{PH}_2 \rightarrow \text{PH} + \text{H}_2$) tends to infinity, as shown in Figure 3E. This indicates that PH₂ is consumed much faster (*instantaneously* relative to the time granularity of the data) than the rest of the species. Therefore, model 5 can do without PH₂ as a variable and the initial abundance of PH₂ can be added to the initial abundance of PH. Reaction 13 ($\text{H} + \text{PH}_3 \rightarrow \text{PH}_2 + \text{H}_2$) produces PH₂, so it can be merged with reaction 12 to yield $2\text{H} + \text{PH}_3 \rightarrow \text{PH} + 2\text{H}_2$, which becomes new reaction 13. Similarly to the reduction step 4, the rate law is derived from the rate-determining step, thus for the new reaction 13, the rate law is $v = k_{13}[\text{H}][\text{PH}_3]$. In addition, since we are assuming that PH₂ is consumed immediately in reaction 12, reaction

3 ($\text{O} + \text{PH}_2 \rightarrow \text{PO} + \text{H}_2$) is also erased. Figure 4E shows that after these changes are applied, the evolution curves of model 5 remain practically equivalent to those of the original model but with the absence of CP and PH₂.

Similarly to the former two steps, in the reduction step 6 (Figure 3F), the rate coefficient k_{13} of reaction $2\text{H} + \text{PH}_3 \rightarrow \text{PH} + 2\text{H}_2$ approaches infinity. In model 6, the initial abundance of PH₃ can, therefore, be added to the initial abundance of PH, removing variable PH₃. This reaction is not merged with any other reaction because no reaction forms PH₃, thus it is simply removed in model 6. Figure 4F shows that in this new model, PO and PN do not match the original curves for short times, but they do so for times

longer than some 100 years. Interestingly, although PH_3 disappears, the evolution of PH does not resemble that of the original PH and, on the contrary, plays a similar role to the original PH_3 . At this stage of the dimensional reduction progression, the chemistry exerted by PH, PH_2 and PH_3 has been condensed in the evolution of a single species, PH.

The reduction step 7 is described in Figure 3G, where reaction rate coefficient k_9 tends to zero, and therefore reaction 9 ($\text{P} + \text{CN} \rightarrow \text{PN} + \text{C}$) can be eliminated. Figure 4G shows that this leads to a time shift with respect to model 6 in Figure 4F: PH abundance decays over longer timescales in model 7 than in model 6, while PO and PN reach detectable abundances ($\sim 10^{-12}$ relative to H) later in model 7 than in model 6. This represents a significant change in the new model for $t < 10^3$ yrs, although between $t \sim 10^3 - 10^5$ yrs the abundance evolution profiles of P, PO and PN remain indistinguishable from those of the original model. In summary, reaction 9 plays an important role in the dynamics at short times, but its omission does not affect the dynamics of PO and PN within the time frame of interest.

Interestingly, the reduction step 8 shown in Figure 3H involves 3 parameters. The rate coefficients of reactions 4 ($\text{O} + \text{PH} \rightarrow \text{PO} + \text{H}$), 10 ($\text{H} + \text{PH} \rightarrow \text{P} + \text{H}_2$), and 14 ($\text{C} + \text{N} + \text{PH} \rightarrow \text{PN} + \text{C} + \text{H}$) approach infinity at the same speed. These reactions consume PH and transform it into PO, P, and PN, respectively. In the new model, these reactions are instantaneous, and the initial abundance of PH can be distributed into the initial abundances of PO, P, and PN, as shown in Supplementary Section S4. These changes yield model 8 that, as can be seen in Figure 4H, only has P, PO and PN as variables. Notably, at this stage of complexity reduction, the system still predicts PO and PN abundances beyond $t = 10^3$ yrs that are identical with the naked eye to those of the original model.

The last iteration of the FISIR algorithm, reduction step 9, is shown in Figure 3I. Reaction rate coefficient k_1 tends to zero, and consequently reaction 1 ($\text{N} + \text{PO} \rightarrow \text{P} + \text{NO}$) is eliminated. Only reactions 2 ($\text{N} + \text{PO} \rightarrow \text{PN} + \text{O}$) and 6 ($\text{P} + \text{OH} \rightarrow \text{PO} + \text{H}$) remain. This is the first model with visible differences between its PO and PN curves and those of the original model for $t > 10^4$ yrs (Figure 4I). Anyway, while model 9 still predicts final abundances of PO and PN close to the original ones, a new reduction step—as described in Supplementary Section S5 —, yields a system with a unique reaction and PO and PN curves completely different from the original model for all times, clearly warning that the model cannot be simplified beyond model 9.

3.2.2 Predictive power and parameter sloppiness of the reduced models

As shown in the previous Subsection, at least one parameter tends toward an extreme value during each successive navigation through the multi-dimensional parameter space. However, it is important to note that the remaining parameters also undergo minor changes. This occurs because the vector $\nu(\tau_i)$ can point in any direction during each path iteration through the parameter space. Additionally, the values of all parameters may change when the new model is fitted. Figure 5A quantifies this phenomenon by showing the evolution of the set of rate coefficients, \mathbf{k} , throughout all dimensional reduction steps, both after each parameter navigation and after each parameter fitting. Notably, the rate coefficients that persist at step 8 (k_1 , k_2 , and k_6) exhibit minimal variation $\sim 0.3\%$ —

compared to their original values from the KIDA database and the work by García de la Concepción et al. (2021).

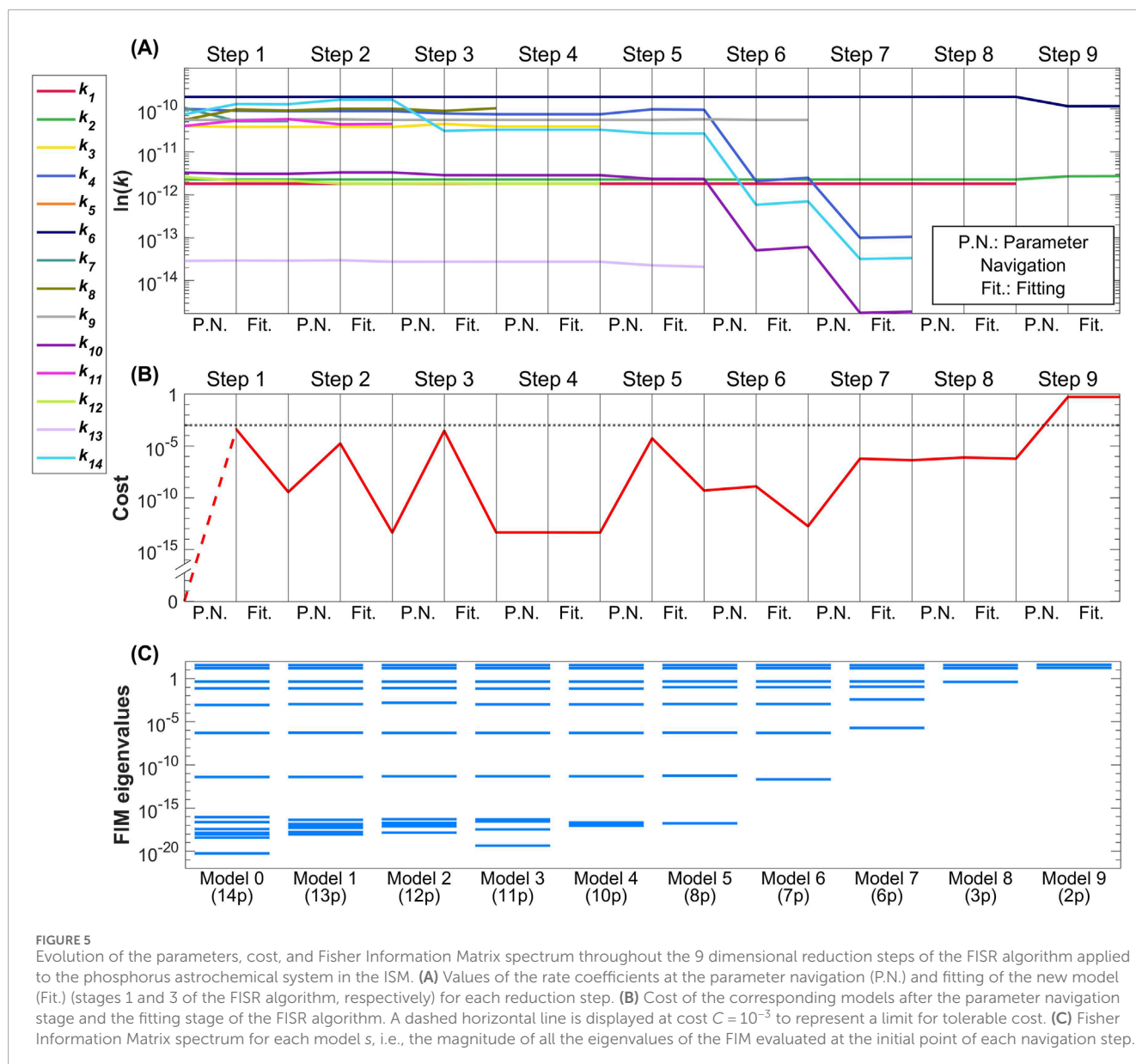
Figure 5B shows the cost—calculated using Equation 2— after each parameter navigation and parameter fitting step. Since the cost associated with model s quantifies the difference between its predictions and those of the original model, low-cost values indicate that the reduction steps preserve the model's predictive power. While the FISIR algorithm identifies the path that minimizes the cost increase, some growth in cost during the parameter navigation stage is inevitable. However, the fitting process consistently offsets this increase to some extent. Consistent with the observation that the abundance plots of PO and PN at long timescales remain indistinguishable from the original curves for models 1–8, the cost during these steps remains very low, reaching 6×10^{-7} at the end of step 8. In contrast, the cost sharply rises to 0.55 by the end of step 9.

Finally, Figure 5C displays the eigenspectrum of each model, i.e., the set of eigenvalues of the Fisher Information Matrix \mathbf{g} evaluated at the initial point $\mathbf{k}(\tau_0)$ of each navigation step through the parameter space. Since each eigenvalue quantifies the model's sensitivity along the direction of its corresponding eigenvector, and the eigenvalues of the original model span 22 orders of magnitude, this plot highlights the significant sloppiness and parameter hierarchy inherent in the current models used to study the chemical evolution of phosphorus in the ISM.

3.3 Final model for the chemical evolution of phosphorus in the ISM

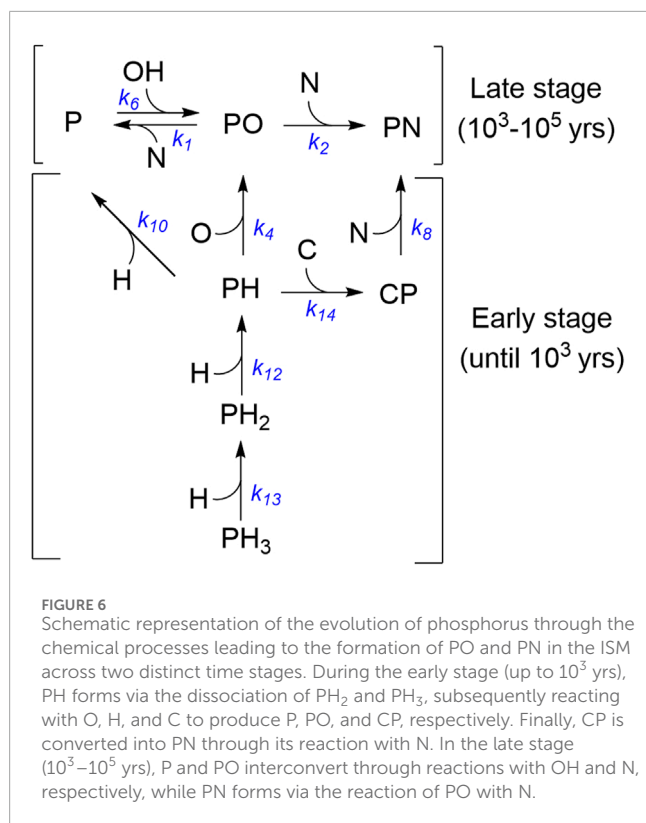
The FISIR method applied in this study enabled the reduction of the original 14-reaction model to significantly simpler models that still accurately reproduce the chemical evolution of PO and PN in the ISM beyond $t \sim 10^3$ yrs—and note that only values beyond 10^4 yrs were used as input in the reduction algorithm. The algorithm was initiated with rate coefficients calculated for a temperature of $T = 100$ K, but note that the results presented here are not restricted to a single temperature but are more general (see the Discussion and Supplementary Section S6 for details). Notably, after eight reduction steps, the chemical kinetics was simplified to include only reactions 1 ($\text{N} + \text{PO} \rightarrow \text{P} + \text{NO}$), 2 ($\text{N} + \text{PO} \rightarrow \text{PN} + \text{O}$), and 6 ($\text{P} + \text{OH} \rightarrow \text{PO} + \text{H}$). Despite this significant reduction, the associated cost ($C \sim 6 \times 10^{-7}$) was so low that the PO and PN abundances beyond $t \sim 10^3$ yrs were indistinguishable from those predicted by the original 14-reaction model. As a result, model 8 provides a simplified and interpretable framework that accurately reproduces the long-term evolution of PO and PN formation predicted by the original model. As illustrated in Figure 4H, P acts as a reservoir due to its much higher abundance than other species. Since P is directly converted into PO through reaction 6, PO becomes more abundant than PN during the early stages of molecular cloud evolution. However, PO is irreversibly transformed into PN via reaction 2, which accounts for the resulting curves where $[\text{PN}] > [\text{PO}]$ for $t > 3 \times 10^4$ yrs.

Model 8 was constructed through eight consecutive reduction steps. Interestingly, while its dynamics is derived from reactions 1, 2, and 6 with rate coefficients almost identical to those in the original model, the overall output is not solely explained by the activity of these reactions. It also incorporates the influence of reactions 4, 8,



10, 12, 13, and 14, all treated as infinitely fast, through modifications in the initial conditions of PH, P, PO, and PN in model 8. These adjustments account for the reactants transformed into products at much faster rates than the remaining relevant reactions. Essentially, these fast processes involved the cascade of $\text{PH}_3 \rightarrow \text{PH}_2 \rightarrow \text{PH}$, the conversion of PH into P, PO, and CP, and the transformation of CP into PN. The elimination of these reactions did not significantly increase the cost, as they occur on much shorter timescales than reactions 1, 2, and 6. Consequently, two distinct stages can be identified in the phosphorus chemical evolution under study: an early stage (up to $t \sim 10^3$ yrs) and a late stage (from $t \sim 10^3$ to 10^5 yrs). Model 8 effectively describes the dynamics of P, PO, and PN during the late stage, which accounts for the majority of the simulation time, while accurately capturing the final outcomes of the early chemistry without the need to reproduce its detailed dynamics. A schematic representation of the processes governing the chemistry in both stages is shown in Figure 6.

With one additional reduction step, reaction 1 is eliminated. Model 9 represents the direct transformation of P into PO and subsequently into PN, but with an associated cost of 0.55 that is 10^5 times the cost of the three-reaction model. This increased cost results in a noticeable discrepancy between the abundance evolution profiles of PO and PN in model 9 and those of the original model. When applying dimensionality reduction techniques, there is always a trade-off between a model's simplicity and the accuracy of its results, with the choice of an appropriate minimal model depending on the specific context and requirements. In this case, model 8 provides highly accurate results (as evidenced by its low cost) while remaining interpretable, making it effective for explaining the formation of PO and PN. However, model 9 offers an even simpler chemical network and may be suitable in situations where a higher cost is acceptable. Finally, further parameter reduction—step 10— leads the system to a single reaction, $\text{N} + \text{PO} \rightarrow \text{PN} + \text{O}$, with a cost that increases to approximately 160.



At this stage, the model can no longer accurately reproduce the abundances of PO and PN, and this last step must be avoided. More information on how to obtain the minimal model with the FISIR method and full details of the reduction process for step 10 are provided in [Supplementary Section S5](#).

4 Discussion

In this work, we introduce a novel dimensionality reduction technique called the Fisher Information Spectral Reduction (FISIR) method. It measures the sensitivity of the output on the parameters in any multi-parameter model, and we used it to reduce the complexity of a system describing the chemical network of phosphorus in the interstellar medium. This P-chemistry model, previously studied in [Fernández-Ruz et al. \(2023\)](#), demonstrates a strong parameter hierarchy, with high dependence on a few parameter directions, while exhibiting a pronounced insensitivity to the rest—commonly referred to as stiff and sloppy directions, respectively. Using the FISIR method, we have exploited this property showing that a kinetic model with 3 chemical reactions ($\text{N} + \text{PO} \rightarrow \text{P} + \text{NO}$, $\text{N} + \text{PO} \rightarrow \text{PN} + \text{O}$, and $\text{P} + \text{OH} \rightarrow \text{PO} + \text{H}$, corresponding to model 8 in [Section 3](#)) and their 3 associated rate coefficients can predict time-evolution abundance profiles of PO and PN in the time range between 10^3 and 10^5 yrs that are indistinguishable—with a cost or error of approximately 10^{-7} —from those of the original model, which consisted of 14 chemical reactions and 14 rate coefficients. Notably, the rate coefficients associated with these 3 chemical reactions remain nearly unchanged throughout the reduction process, suggesting that even subtle changes in these

rate coefficients could drastically affect the system's predictions. In fact, their values differ by less than 0.3% between the original and any of the reduced models, while the uncertainties associated with them in the KIDA database correspond to a 1σ confidence interval bounded by $k/2$ and $2k$ ([Wakelam et al., 2012](#)). In consequence, as a key takeaway for observational and theoretical astrochemists, our findings emphasize the importance of distinguishing between sloppy and stiff parameters in astrochemical models to target the key chemical reactions (the stiff parameters) whose rate coefficients need to be determined with high accuracy either via laboratory experiments or quantum chemical computations. A good example is the rate coefficient of the reaction $\text{P} + \text{OH} \rightarrow \text{PO} + \text{H}$, found to be essential in our model and also in reproducing the observed $[\text{PO}]/[\text{PN}]$ ratios ([Jiménez-Serra et al., 2018](#)), and which has recently been determined by quantum chemical calculations ([García de la Concepción et al., 2021](#)). Other works have recently explored the rate coefficients of the reactions $\text{N} + \text{PO} \rightarrow \text{P} + \text{NO}$ and $\text{N} + \text{PO} \rightarrow \text{PN} + \text{O}$ ([Douglas et al., 2022](#)), although the level of theory of these quantum chemical computations is low.

We began this article by outlining the various challenges that phosphorus astrochemistry and its connection to the origin of life on Earth still present. Let us now examine how the drastic simplification of the phosphorus astrochemical network to its fundamental core, as presented here, enables a deeper understanding of its chemistry. Despite the minimal model 8 having only 3 rate coefficients as inputs, it accounts for the effects of other reactions without explicitly including their rate coefficients in the equations. The reduction steps in which one or more rate coefficients tend to infinity are interpreted as the corresponding reactions being instantaneous, and this translates into appropriate changes in the initial conditions. As shown in [Figure 4](#), in the original model (dashed lines), by $t \sim 10^3$ yrs all CP and PH are already depleted, in part due to the dehydrogenation of PH_2 and PH_3 . Therefore, the velocities of these processes are irrelevant for the long-term dynamics of P, PO, and PN as long as the associated reactions have reached equilibrium by $t = 10^3$ yrs. Based on this, we were able to explain PO and PN formation as a two-step process: (i) an early stage, that can be considered infinitely fast; and (ii) a late stage, whose dynamics must be explicitly modeled with reactions 1, 2 and 6. The 2-stage interpretation of P-chemistry in the ISM implies that the exact values of the rate coefficients for the reactions operating on short timescales are not necessary to accurately predict the abundance evolution profiles of PO and PN on long timescales. However, the counterpart is that nothing about the early chemistry (involving PH, PH_2 , PH_3 and CP) can be inferred solely from the PO and PN detections, since we demonstrated that the observed abundances of PO and PN do not serve to constrain the values of the rate coefficients of the early stage. A more complex description of P-chemistry on short timescales in star-forming regions would require a more complete set of observational data, including, for example, the abundance of PH_3 in the interstellar medium (especially toward the same regions where PO and PN have been detected). PH_3 , however, remains elusive ([Agúndez et al., 2008](#)). Recent high-sensitivity observations carried out with ALMA toward dark molecular clouds have not yielded any detection (upper limit to the PH_3 abundance of $\leq 5 \times 10^{-12}$; [Rivilla et al. \(2018\)](#), [Furuya and Shimonishi \(2024\)](#)), which may be due to: i) PH_3 not being sufficiently abundant in the gas phase (e.g., it may be heavily frozen-out onto ices

or converted very rapidly into other P-bearing molecules), or ii) PH_3 being very difficult to observe from the ground given that its lowest energy transitions appear at sub-millimeter wavelengths (Agúndez et al., 2008).

It is worth noting that, although the FISR algorithm was applied using rate coefficients calculated for $T = 100$ K, the resulting reduced models remain good approximations (i.e., with low associated cost) over a certain temperature range. This is due to the system's inherent sloppiness, which confers robustness to variations in the rate coefficients and, consequently, to changes in temperature. In [Supplementary Section S6](#), we analyze the generalization of models 1–9—originally obtained at $T = 100$ K—to temperatures ranging from 40 K to 160 K, without reapplying the full algorithm at each temperature. This robustness is advantageous, as it removes the need to re-run the computationally expensive FISR algorithm for every temperature. We find that the dynamics of phosphorus chemistry in the 40–160 K range can still be interpreted as a two-stage process, being the approximation of the early stage as instantaneous particularly accurate at higher temperatures. At very low temperatures, applying the full FISR method becomes necessary to obtain precise predictions. Accordingly, [Supplementary Section S8](#) shows the costs of models 1–9 at $T = 40, 70, 100, 130$, and 160 K.

Relative to other dimensionality reduction methods currently available in the literature, the FISR method proposed here is significantly simpler. Specifically, in comparison to the MBAM algorithm (Transtrum and Qiu, 2014), upon which it is based, the FISR algorithm employs a streamlined methodology that is easier to implement, conceptually simpler, and admits larger output vectors (see [Supplementary Section S7](#)). Nevertheless, it is important to note that neither the FISR method nor the MBAM is currently suitable for application to very large systems due to several limitations: (i) both are based on a streamlined algorithmic structure that cannot be parallelized, (ii) each point of the trajectory navigating the parameter space requires the diagonalization of the Fisher Information Matrix computed numerically from the solution of the ODEs—in the phosphorus network, we were able to solve the model analytically, but this will not generally be feasible—and (iii) manual analysis of each reduced model is required at the end of each reduction step. In summary, it is crucial to continue advancing this promising line of research by developing new algorithms capable of analyzing more complex systems. Such advancements will enable a comprehensive study of the chemistry of the interstellar medium, where we believe parameter sloppiness is the rule rather than the exception.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Author contributions

MF-R: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review

and editing, Data curation, Visualization. IJ-S: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing. MC: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing, Formal Analysis, Software. MR-B: Investigation, Writing – original draft, Writing – review and editing, Visualization. JA: Investigation, Writing – original draft, Writing – review and editing, Conceptualization, Formal Analysis, Funding acquisition, Methodology, Project administration, Software, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. JA and MF-R received support from grant No. PID2021-122936NB-I00, MC from grant No. PID2022-140217NB-I00, MR-B from grant PID2022-140180OB-C22, and IJ-S from grant PID2022-136814NB-I00, all of them funded by the Spanish Ministry of Science and Innovation/State Agency of Research MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”. JA acknowledges financial support from grant No. PIE2024ICT085 and IJ-S from grant No. PIE 202250E155, both of them funded by the Spanish National Research Council (CSIC). MR-B is supported by the Instituto Nacional de Técnica Aeroespacial “Esteban Terradas” (INTA). MF-R is supported by a predoctoral contract from INTA. Authors benefited from the interdisciplinary framework provided by CSIC through the “LifeHUB.CSIC” initiative (PIE 202120E047-Conexiones-Life). IJ-S also acknowledges funding from the ERC grant OPENS (project number 101125858) funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgments

The authors are indebted to S. Ares, J. A. Cuesta and M. Martínez-Jiménez for fruitful comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FF declared a past co-authorship with the author(s) IJ-S to the handling editor.

Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript exclusively to

review the English precision of some sentences in the manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim

that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspas.2025.1570525/full#supplementary-material>

References

- Agúndez, M., Cernicharo, J., and Guélin, M. (2007). Discovery of phosphathethyne (HCP) in space: phosphorus chemistry in circumstellar envelopes. *Astrophysical J. Lett.* 662, L91–L94. doi:10.1086/519561
- Agúndez, M., Cernicharo, J., Pardo, J. R., Guélin, M., and Phillips, T. G. (2008). Tentative detection of phosphine in IRC +10216. *Astronomy Astrophysics* 485, L33–L36. doi:10.1051/0004-6361/200810193
- Agúndez, M., and Wakelam, V. (2013). Chemistry of dark clouds: databases, networks, and models. *ChRv* 113, 8710–8737. doi:10.1021/cr4001176
- Antoulas, A. C. (2005). *Approximation of large-scale dynamical systems*. Philadelphia: SIAM.
- Aota, T., and Aikawa, Y. (2012). Phosphorus chemistry in the shocked region L1157 B1. *Astrophysical J.* 761, 74. doi:10.1088/0004-637X/761/1/74
- Bernal, J. J., Koelmay, L. A., and Ziurys, L. M. (2021). Detection of PO in orion-KL: phosphorus chemistry in the plateau outflow. *Astrophysical J.* 906, 55. doi:10.3847/1538-4357/abc87b
- Brown, K. S., and Sethna, J. P. (2003). Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* 68, 021904. doi:10.1103/physreve.68.021904
- Chantzios, J., Rivilla, V. M., Vasyunin, A., Redaelli, E., Bizzocchi, L., Fontani, F., et al. (2020). The first steps of interstellar phosphorus chemistry. *Astronomy Astrophysics* 633, A54. doi:10.1051/0004-6361/201936531
- Charnley, S. B., and Millar, T. J. (1994). The chemistry of phosphorus in hot molecular cores. *MNRAS* 270, 570–574. doi:10.1093/mnras/270.3.570
- Cooper, G. W., Onwo, W. M., and Cronin, J. R. (1992). Alkyl phosphonic acids and sulfonic acids in the murchison meteorite. *Geochimica Cosmochimica Acta* 56, 4109–4115. doi:10.1016/0016-7037(92)90023-C
- de la Concepción, J. G., Cavallotti, C., Barone, V., Puzzarini, C., and Jiménez-Serra, I. (2024). Relevance of the P+O₂ reaction for PO formation in astrochemical environments: electronic structure calculations and kinetic simulations. *Astrophysical J.* 963, 142. doi:10.3847/1538-4357/ad1ffa
- Douglas, K. M., Gobrecht, D., and Plane, J. M. C. (2022). Experimental study of the removal of excited state phosphorus atoms by H₂O and H₂: implications for the formation of PO in stellar winds. *Mon. Notices R. Astronomical Soc.* 515, 99–109. doi:10.1093/mnras/stac1684
- Faro, J., Von Haften, B., Gardner, R., and Faro, E. (2019). A sensitivity analysis comparison of three models for the dynamics of germinal centers. *Front. Immunol.* 10, 2038. doi:10.3389/fimmu.2019.02038
- Fernández-Ruz, M., Jiménez-Serra, I., and Aguirre, J. (2023). A theoretical approach to the complex chemical evolution of phosphorus in the interstellar medium. *Astrophysical J.* 956, 47. doi:10.3847/1538-4357/acf290
- Fontani, F. (2024). Observations of phosphorus-bearing molecules in the interstellar medium. *Front. Astronomy Space Sci.* 11, 1451127. doi:10.3389/fspas.2024.1451127
- Fontani, F., Mininni, C., Beltrán, M. T., Rivilla, V. M., Colzi, L., Jiménez-Serra, I., et al. (2024). The GUAPOS project: G31.41+0.31 Unbiased ALMA sPectral Observational Survey. IV. Phosphorus-bearing molecules and their relation to shock tracers. *Astronomy Astrophysics* 682, A74. doi:10.1051/0004-6361/202348219
- Fontani, F., Rivilla, V. M., Caselli, P., Vasyunin, A., and Palau, A. (2016). Phosphorus-bearing molecules in massive dense cores. *Astrophysical J. Lett.* 822, L30. doi:10.3847/2041-8205/822/2/L30
- Furuya, K., and Shimonishi, T. (2024). Deep search for phosphine in a prestellar core. *Astrophysical J. Lett.* 968, L19. doi:10.3847/2041-8213/ad50cc
- García de la Concepción, J., Puzzarini, C., Barone, V., Jimenez-Serra, I., and Roncero, O. (2021). Formation of phosphorus monoxide (PO) in the interstellar medium: insights from quantum-chemical and kinetic calculations. *Astrophysical J.* 922, 169. doi:10.3847/1538-4357/ac1e94
- Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L-error bounds. *Int. J. Control* 39, 1115–1193. doi:10.1080/00207178408933239
- Goldsmith, P. F., Liseau, R., Bell, T. A., Black, J. H., Chen, J.-H., Hollenbach, D., et al. (2011). Herschel measurements of molecular oxygen in Orion. *Astrophysical J.* 737, 96. doi:10.1088/0004-637x/737/2/96
- Gomes, A. C. R., Souza, A. C., Jasper, A. W., and Galvão, B. R. L. (2023). The P(⁴S) + NH(³Σ⁻) and N(⁴S) + PH(³Σ⁻) reactions as sources of interstellar phosphorus nitride. *Astronomical Soc. Aust.* 40, e011. doi:10.1017/pasa.2023.13
- Guélin, M., Cernicharo, J., Paubert, G., and Turner, B. E. (1990). Free CP in IRC +10216. *Astronomy Astrophysics* 230, L9–L11.
- Gugercin, S., and Antoulas, A. C. (2004). A survey of model reduction by balanced truncation and some new results. *Int. J. Control* 77, 748–766. doi:10.1080/00207170410001713448
- Halfen, D. T., Clouthier, D. J., and Ziurys, L. M. (2008). Detection of the CCP radical (X²Π_g) in IRC +10216: a new interstellar phosphorus-containing species. *Astrophysical J.* 677, L101–L104. doi:10.1086/588024
- Holdship, J., Viti, S., Jiménez-Serra, I., Makrymalis, A., and Priestley, F. (2017). UCLCHEM: a gas-grain chemical code for clouds, cores, and C-shocks. *Astronomical J.* 154, 38. doi:10.3847/1538-3881/aa773f
- Jiménez-Serra, I., Viti, S., Quénard, D., and Holdship, J. (2018). The chemistry of phosphorus-bearing molecules under energetic phenomena. *Astrophysical J.* 862, 128. doi:10.3847/1538-4357/aacd2f
- Lagarias, J., Reeds, J., Wright, M., and Wright, P. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM J. Optim.* 9, 112–147. doi:10.1137/S1052623496303470
- Larsson, B., Liseau, R., Pagani, L., Bergman, P., Bernath, P., Biver, N., et al. (2007). Molecular oxygen in the ρ Ophiuchi cloud. *A&A* 466, 999–1003. doi:10.1051/0004-6361:20065500
- Lefloch, B., Codella, C., Montargès, M., Vastel, C., Podio, L., Viti, S., et al. (2024). Spatial distributions of PN and PO in the shock region L1157-B1. *Astronomy Astrophysics* 687, A75. doi:10.1051/0004-6361/202245338
- Lefloch, B., Vastel, C., Viti, S., Jimenez-Serra, I., Codella, C., Podio, L., et al. (2016). Phosphorus-bearing molecules in solar-type star-forming regions: first PO detection. *MNRAS* 462, 3937–3944. doi:10.1093/mnras/stw1918
- Maciá-Barber, E. (2020). *The chemical evolution of phosphorus: an interdisciplinary approach to astrobiology*. Oakville, Ontario, Canada: Apple Academic Press.
- Moore, B. C. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automatic Control* 26, 17–32. doi:10.1109/tac.1981.1102568
- Pilorget, C., Baklouti, D., Bibring, J.-P., Brunetto, R., Ito, M., Franchi, I., et al. (2024). Phosphorus-rich grains in Ryugu samples with major biochemical potential. *Nat. Astron.* 8, 1529–1535. doi:10.1038/s41550-024-02366-w
- Quinn, K. N., Abbott, M. C., Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2023). Information geometry for multiparameter models: new perspectives on the origin of simplicity. *Rep. Prog. Phys.* 86, 035901. doi:10.1088/1361-6633/aca6f8
- Rivilla, V. M., Drozdovskaya, M. N., Altwegg, K., Caselli, P., Beltrán, M. T., Fontani, F., et al. (2020). ALMA and ROSINA detections of phosphorus-bearing molecules: the interstellar thread between star-forming regions and comets. *MNRAS* 492, 1180–1198. doi:10.1093/mnras/stz3336
- Rivilla, V. M., Fontani, F., Beltrán, M. T., Vasyunin, A., Caselli, P., Martín-Pintado, J., et al. (2016). The first detections of the key prebiotic molecule PO in star-forming regions. *Astrophysical J.* 826, 161. doi:10.3847/0004-637X/826/2/161
- Rivilla, V. M., García De La Concepción, J., Jiménez-Serra, I., Martín-Pintado, J., Colzi, L., Tercero, B., et al. (2022). Ionize hard: interstellar PO+ detection. *Front. Astronomy Space Sci.* 9, 829288. doi:10.3389/fspas.2022.829288

- Rivilla, V. M., Jiménez-Serra, I., Zeng, S., Martín, S., Martín-Pintado, J., Armijos-Abendaño, J., et al. (2018). Phosphorus-bearing molecules in the galactic center. *MNRAS Lett.* 475, L30–L34. doi:10.1093/mnrasl/slx208
- Ruaud, M., Wakelam, V., and Hersant, F. (2016). Gas and grain chemical composition in cold cores as predicted by the Nautilus three-phase model. *Mon. Notices R. Astronomical Soc.* 459, 3756–3767. doi:10.1093/mnras/stw887
- Rugel, M. R., Beuther, H., Bihr, S., Wang, Y., Ott, J., Brunthaler, A., et al. (2018). OH absorption in the first quadrant of the Milky Way as seen by THOR. *Astronomy Astrophysics* 618, A159. doi:10.1051/0004-6361/201731872
- Saltelli, A. (2008). *Global sensitivity analysis: the primer*. John Wiley and Sons.
- Scibelli, S., Megías, A., Jiménez-Serra, I., Shirley, Y., Bergner, J., Ferrer Asensio, J., et al. (2025). First detections of PN, PO, and PO+ toward a shocked low-mass starless core. *Astrophysical J. Lett.* 985, L25. doi:10.3847/2041-8213/add344
- Sil, M., Srivastav, S., Bhat, B., Mondal, S. K., Gorai, P., Ghosh, R., et al. (2021). Chemical complexity of phosphorous-bearing species in various regions of the interstellar medium. *Astronomical J.* 162, 119. doi:10.3847/1538-3881/ac09f9
- Tenenbaum, E. D., Woolf, N. J., and Ziurys, L. M. (2007). Identification of phosphorus monoxide ($X^2\Pi_r$) in VY Canis majoris: detection of the first P-O bond in space. *Astrophysical J.* 666, L29–L32. doi:10.1086/521361
- Tenenbaum, E. D., and Ziurys, L. M. (2008). A search for phosphine in circumstellar envelopes: PH_3 in IRC +10216 and CRL 2688? *Astrophysical J.* 680, L121–L124. doi:10.1086/589973
- Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2011). Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E* 83, 036701. doi:10.1103/PhysRevE.83.036701
- Transtrum, M. K., and Qiu, P. (2014). Model reduction by manifold boundaries. *Phys. Rev. Lett.* 113, 098701. doi:10.1103/PhysRevLett.113.098701
- Transtrum, M. K., and Qiu, P. (2016). Bridging mechanistic and phenomenological models of complex biological systems. *PLOS Comput. Biol.* 12, e1004915. doi:10.1371/journal.pcbi.1004915
- Turner, B. E., and Bally, J. (1987). Detection of interstellar PN: the first identified phosphorus compound in the interstellar medium. *Astrophysical J. Lett.* 321, L75. doi:10.1086/185009
- Wakelam, V., Herbst, E., Loison, J.-C., Smith, I. W. M., Chandrasekaran, V., Pavone, B., et al. (2012). A kinetic database for astrochemistry (KIDA). *Astrophysical J. Suppl. Ser.* 199, 21. doi:10.1088/0067-0049/199/1/21
- Walton, C., Ewens, S., Coates, J., Blake, R., Planavsky, N., Reinhard, C., et al. (2023). Phosphorus availability on the early earth and the impacts of life. *Nat. Geosci.* 16, 399–409. doi:10.1038/s41561-023-01167-6
- Ziurys, L. M. (1987). Detection of interstellar PN: the first phosphorus-bearing species observed in molecular clouds. *Astrophysical J. Lett.* 321, L81. doi:10.1086/185010