



Institute of Chemistry  
Postgraduate Program in Chemistry  
Doctoral Thesis

**KINETICS OF ENVIRONMENTAL BIOCOMPLEXITY:  
EXPERIMENTS, QUANTUM CHEMISTRY AND MACHINE  
LEARNING**

Author: Flávio Olímpio Sanches Neto

Advisor: Dr. Valter Henrique Carvalho-Silva

Co-advisor: Dr. Kleber Carlos Mundim

Brasília, DF

2022



**UnB**  
Doctoral Thesis

---

**KINETICS OF ENVIRONMENTAL BIOCOMPLEXITY:  
EXPERIMENTS, QUANTUM CHEMISTRY AND MACHINE  
LEARNING**

---

Author: Flávio Olímpio Sanches Neto  
Advisor: Dr. Valter Henrique Carvalho-Silva  
Co-advisor: Dr. Kleber Carlos Mundim

A thesis submitted in fulfillment of the  
requirements for the degree of Doctor of  
Philosophy. Institute of Chemistry

Brasília, DF

2022

## ACKNOWLEDGMENTS

First of all, I would like to thank my mother, Eunice de Souza Gonçalves, a warrior woman and who always gave me full support to be able to get an academic education, without her help it would certainly not have been possible to reach this point. In the person of my mother, I also thank all my family for their motivation and for believing that this dream was possible.

To the friends of the Graduate Program in Chemistry at UnB, UEG, UFG, UFSCAR and USP that I made during this journey, for all the jokes and fun moments that made this journey happy. In particular, I thank my friends: Hugo Gontijo Machado, Jefferson Richard, Vitor Gomes, Priscila Gomes, Lilian Carvalho, Allane Carvalho, Ingrid Vieira, Leonardo Almeida, Anderson Catão, Herika Gomes. And to the wronged, people who helped me during this period and who I forgot to mention here.

I would also like to thank my co-workers at the Instituto Federal de Goiás – Anápolis campus, for having helped me richly with the schedules to be able to reconcile work and my doctorate – especially Eder's help. And on behalf of Professor Elza, director of the Institute, I would like to express my gratitude to all my colleagues. I also want to thank my work colleagues at the Federal University of Goiás, who have helped me a lot in recent months. In the person of the Director of the Institute of Chemistry, Wendell Coltro, I thank all the other colleagues. I also want to highlight the thanks to the laboratory technician Juvan, who helped me a lot in the experimental classes with his rich wisdom.

I thank Professor Vinczo Aquilanti, who made the technical visit possible at the University of Perugia in Italy, and for his rich contributions to this thesis. It is a pleasure to be able to discuss high-level science with a prestigious person on the world scientific scene.

To my supervisor Professor Dr. Valter Henrique Carvalho-Silva, who I had the opportunity to work with since graduation and who contributed richly to my training, it was an honor to have him as a supervisor. I appreciate his tips and

the “network” he provided me with high-level researchers. I learned to always work with dedication and professionalism, things that I will take with me for life. I hope that we can continue with this friendship and that great works are produced.

To the professors I had during my doctorate, especially Professor Kleber Carlo Mundim, who is also my co-supervisor and who contributed significantly to the discussions and proposals of this thesis. Your scientific thinking is brilliant and makes us excited to enrich our knowledge. I also thank the other teachers, such as Professor Heibbe Cristian and João Batista.

The collaborations that were carried out for this work to be done. In particular, I highlight the collaborations with Bruno Ramos, which made possible a partnership that ended up being consolidated in a collaboration with the University of São Paulo and with the approval of a project financing between FAPEG and FAPESP and for his contributions during my qualification. In the same sense, I thank Professor Antonio Carlos and Larissa for their collaboration and their patience in helping me on the technical visit to USP. I also want to richly thank the contributions of Jefferson Richard who supported me, and, because of this partnership, we were able to obtain good works already published and with certain future partnerships. With his collaboration I also had the opportunity to meet his supervisor Luiz Keng and Vitor Gomes, whom I had the opportunity to collaborate with and for that I would like to express my gratitude for the rich discussions and contributions in this thesis.

Throughout my doctorate I had the opportunity to make good relationships which resulted in some collaborations. I highlight the collaboration with Professor Tiago Alves, whom I met at SBQT in João Pessoa, and we have built some projects in common. Some relationships with people outside the public initiative were carried out, and here I highlight the partnership with Jefferson Luiz Fonseca, from which in recent years we have developed projects related to optimization processes in the pharmaceutical industry, mainly in analytical development.

I also want to highlight the various contributions, whether in a conversation over a coffee or in a meeting I had with the professors of the QTEA group, especially professors Ademir Carmargo, Luciano Ribeiro and Hamilton Napolitano.

I thank the examining board, Professor Ricardo Gargano, Ademir Camargo and Daniel Scalabrini, for agreeing to share this moment.

I would like to thank the PPGQ doctoral secretariats at the University of Brasília.

I thank the funding institutions, CAPES, FAPDF, FAPEG and the University of Brasília for the incentive and financial investment provided for this work to be carried out.

## DEDICATION

*I dedicate this work to my dear grandmother, Nerica Fernandez de Paula, and my beloved father Adverso Olimpio Sanches. The longing and the memories encourage me to always go on and do my best.*

## ABSTRACT

Micro-pollutants of emerging concern have imposed a major technological challenge: pesticides, drugs and other anthropogenic substances are increasingly found in aquatic and atmospheric environments and even in water supplies, being related to adverse effects on biota and human health. Overcoming this challenge requires understanding the behavior of these species in the environment and the development of technologies that allows for minimizing their dissemination. Viable alternatives applied in this thesis include the use of radical-based oxidation processes using both experimental – via the competition kinetics method – and theoretical protocols – blend of kinetic, quantum chemistry and machine learning calculations. In a first study, the mechanisms, kinetics, and an evaluation of the toxicity of picloram degradation – a pesticide widely used in the world - initiated by OH radicals indicate that: i) two favorable pathways occur by addition to the pyridine ring, ii) picloram and the majority of degradation products are estimated as harmful; however, ii) these compounds can suffer photolysis by sunlight. However, the competition kinetic method and the quantum chemistry description make the degradation analyses a formidable enterprise, considering the costs of ad hoc instrumental equipment's and dedicated computational efforts. To overcome the demanding conventional procedures, we developed a free and user-friendly web application ([www.pysirc.com.br](http://www.pysirc.com.br)) based on holistic machine learning combined with molecular fingerprints models that permits compilation of kinetic parameters and mechanistic interpretation of radical-based oxidation attacks according to the OECD principles. Machine learning algorithms were implemented, and all models provided high goodness-of-fit for radical-based degradation in aquatic and atmospheric environment. The models were interpreted using the SHAP (SHapley Additive exPlanations) method: the results showed that the model developed made the prediction based on a reasonable understanding of how electron-withdrawing/donating groups interfere in the reactivity of the radicals. We argue that our models and web interface can stimulate and expand the application and interpretation of kinetic research on contaminants in water and air treatment units based on advanced oxidative technologies.

## RESUMO

Micro poluentes de preocupação emergente têm imposto um grande desafio tecnológico: pesticidas, drogas e outras substâncias antropogênicas são cada vez mais encontrados em ambientes aquáticos e atmosféricos e até mesmo no abastecimento de água, estando relacionados a efeitos adversos sobre a biota e a saúde humana. Superar esse desafio requer a compreensão do comportamento dessas espécies no meio ambiente e o desenvolvimento de tecnologias que permitam minimizar sua disseminação. Alternativas viáveis aplicadas nesta tese incluem o uso de processos de oxidação baseado em radicais utilizando tanto o método experimental – através do método cinético de competição – quanto os protocolos teóricos – um conjunto de cálculos cinéticos, quânticos e aprendizado de máquina. Em um primeiro estudo, os mecanismos, cinéticas e uma avaliação da toxicidade da degradação do picloram – pesticida amplamente utilizado no mundo – iniciados por radicais OH indicam que: i) duas vias favoráveis ocorrem por adição ao anel de piridina, ii) picloram e a maioria dos produtos de degradação são estimados como prejudiciais; no entanto, ii) esses compostos podem sofrer fotólise pela luz solar. No entanto, o método cinético da competição e a descrição da química quântica fazem da degradação uma empreendimento formidável, considerando os custos de equipamentos instrumentais *ad hoc* e esforços computacionais dedicados. Para superar os exigentes procedimentos convencionais, desenvolvemos uma aplicação web gratuita e de fácil acesso ([www.pysirc.com.br](http://www.pysirc.com.br)) baseada no aprendizado de máquina holístico combinado com modelos de impressões digitais moleculares que permitem a compilação de parâmetros cinéticos e interpretação mecanicista de ataques de oxidação baseado em radicais de acordo com os princípios da OCDE. Algoritmos de aprendizagem de máquina foram implementados, e todos os modelos forneceram alto desempenho de ajuste para a degradação baseado em radical no ambiente aquático e atmosférico. Os modelos foram interpretados utilizando o método SHAP (Explicações Aditivas de SHapley): os resultados mostraram que o modelo desenvolvido fez a previsão com base em uma compreensão razoável de como grupos de retirada/doação de elétrons interferem na reatividade dos radicais. Argumentamos que nossos modelos e interface web podem estimular e expandir a aplicação e interpretação de pesquisas cinéticas sobre contaminantes em unidades de tratamento de água e ar com base em tecnologias oxidativas avançadas.

## LIST OF FIGURES

<b>Figure 1:</b> Graphical representation of the Arrhenius plot, where the intercept on the ordinate axis indicates the logarithm of the pre-exponential factor and the tangent of the line provides the ratio of the activation energy by the universal gas constant. ....	16
<b>Figure 2:</b> Visualization of the potential energy surface for a chemical reaction with a two-dimensional and (b) three-dimensional representation with the main topographic features. Figure 2b has been adapted from reference 1. ....	17
<b>Figure 3:</b> Representative scheme of trajectories that cross the saddle point. The solid line indicates the case of the transition state located in a fixed position, while the dashed line indicates a variational transition state. ....	22
<b>Figure 4:</b> Eckart potential barrier. ....	25
<b>Figure 5:</b> Arrhenius diagram illustrating tunneling regimes: negligible, moderate and deep. The point where the red and purple dotted line intersect is the crossover temperature, $T_c$ . Figure taken from reference 40. ....	27
<b>Figure 6:</b> An overview for developing a machine learning model. ....	30
<b>Figure 7:</b> Supervised machine learning types. ....	31
<b>Figure 8:</b> Architecture of a decision tree-based algorithm. ....	32
<b>Figure 9:</b> Processing a neural network. ....	33

## LIST OF TABLES

Table 1: Contribution types associated with energy and the partition function.....	19
Table 2: Values of wavelengths of different particles with their respective masses, considering a kinetic energy of $15 \text{ kcal mol}^{-1}$ ( $62.76 \text{ kJ mol}^{-1}$ ).....	23

## CONTENTS

1. INTRODUCTION	12
2. THEORETICAL FOUNDATIONS	14
2.1 Chemical Kinetics	14
2.2 Conventional Transition State Theory	17
2.2.1 Transition State Theory in Thermodynamic Formalism	20
2.3 Corrections to the Transition State Theory	20
2.3.1 Variational Transition State Theory	21
2.3.2 Tunneling corrections	23
2.4 Machine Learning Fundamentals	28
3. PAPERS	34
5. CONCLUSIONS AND PERSPECTIVES	74
6. REFERENCES	75

# 1. INTRODUCTION

The development of chemical kinetics as a fundamental science culminates with the famous Arrhenius equation, that correlates the rate constant with the reciprocal of temperature<sup>1-3</sup>. In the following years, with the advent of quantum mechanics and statistical thermodynamics, Eyring proposed from ab initio concept interpretations to the phenomenological model proposed by Arrhenius: the celebrated Transition State Theory (TST)<sup>4-6</sup>. Although these equations are still widely used – due to the simplicity and accuracy of the results for a set of chemical reactions – the improvement of experimental techniques allowed the assessment of the rate constant in a wide temperature range and the observation of marked deviations in the previous formulations<sup>1,7</sup>. Because of this, several proposals have been developed to include corrections in TST, such as variational, quantum and solvent effects<sup>8-11</sup>. However, even with more recent proposals to account for different effects that affect the kinetics of a reaction, there are other bottlenecks for obtaining the rate constant, namely, accurate calculations of electronic structure<sup>12,13</sup>. From a computational point of view, it is very costly to calculate stationary points with the increase in the number of atoms, thus limiting high-precision results to reactions involving few atoms<sup>14,15</sup>. In addition to the experimental and theoretical approaches to estimate the rate constant of chemical reactions, in recent decades in-silico methodologies have gained attention to make the prediction of the rate constant from molecular descriptors<sup>16-19</sup>.

Recently, data-driven analysis combined with Machine Learning (ML) algorithms leveraged research related to chemical properties<sup>20-22</sup>, including the rate constant<sup>23,24</sup>. Within this area, several protocols have been developed to predict kinetic parameters with precision comparable to experimental and theoretical approaches<sup>25,26</sup>. In this scenario, the quality and quantity of data related to the property to be predicted are essential and are considered one of the greatest challenges for the construction of reliable modlinks<sup>27,28</sup>. To overcome this problem, some studies used a gas phase reaction data set to predict adequate temperatures<sup>26</sup> and chemical reaction activation energies<sup>25,29</sup>; although these data are generally non-homogeneous and do not include the rate constant<sup>30</sup>. Bowman et al<sup>23</sup> used an ML approach – a Gaussian process regression – to predict rate constants over a wide temperature range and provided an 80% accurate response. Komp and Valleau<sup>30</sup> used a deep neural network to predict the product of the logarithm of the rate constant with the partition function and its model presented a relative error of 1.1%, however, for reactions below 300K an error of 31% was observed for the exact results. In recent years, Zhang's group<sup>24,31,32</sup> developed predictive models of the photodegradation and degradation of organic pollutants in an aqueous environment and obtained good accuracy when compared to the literature.

Understanding the rate constant is considered one of the central points in chemistry, since to understand and control reactional processes, information about this parameter and related variables is needed<sup>1,4,33</sup>. In the environmental area, experimental, theoretical and in silico studies have been established to predict the rate constant of oxidation-based reactions in an aqueous environment of organic contaminants<sup>34,35</sup>. Organic contaminants (OCs) are natural or

anthropogenic substances from different human activities<sup>36,37</sup>. The presence of these compounds or metabolites in water bodies or in the atmosphere can have an environmental impact and/or cause climate change and even affect human health when they remain for a long period in the environment<sup>38,39</sup>. The mineralization or degradation processes of these compounds consist of physical, chemical and biological steps<sup>36,40</sup>. However, most of these conventional processes are not considered efficient for the complete removal of these pollutants.<sup>41,42</sup> On the other hand, it is known that only the removal of the starting products is not enough, since, in general, the degradation products are more stable and more polar than the original compound, providing a greater risk to the environment.<sup>43-45</sup> Therefore, detailed knowledge of the mechanism, kinetics and toxicity of COs are considered key steps for a sustainable application of a mitigation of these compounds.<sup>43,46,47</sup> The proposals of this work set out in this direction, they are: (i) revealing the mechanism, kinetics and toxicity of the pesticide picloram with the hydroxyl radical, and (ii) developing ML protocols to predict the rate constant of organic pollutants based on oxidative processes, combined with the development of a free and easily accessible web platform to help the scientific community and technical managers.

## 2. THEORETICAL FOUNDATIONS

### 2.1 Chemical Kinetics

The genesis of chemical kinetics appears late in the mid-eighteenth century with the questioning of the variable time in chemical processes<sup>4,48,49</sup>. The understanding of the reaction mechanisms and the rate of chemical reactions has become, therefore, one of the principles in chemistry. However, the complexity of interpreting such phenomena has made it difficult to advance studies in this area, as Formosinho reports<sup>48</sup> in his book: *"It was the difficulty in finding simplicity in macroscopic observation and in the corresponding microscopic interpretation that delayed the development of Chemical Kinetics."*<sup>50,51</sup>, in 1850, entitled "The Law of Action of Acid in Sugarcane". In this work, he studied the progress of the sugarcane reaction in the presence of acids and showed that:

$$v = -k[A][B], \quad (1)$$

where  $v$  is the reaction rate,  $k$  is the rate constant, and  $[A]$  and  $[B]$  are the sugar and acid concentrations, respectively.

However, this remarkable work on the rate of chemical reactions went unnoticed for a few years. Other studies, such as that of Berthelot and St. Gilles<sup>52</sup>, in the esterification reactions, concluded that the amount of ester as a function of time was proportional to the product of the "active masses", and Guldberg and Waage<sup>53,54</sup>, in the formulation of the "law of the actions of the masses" that was achieved due to the introduction of the concept of chemical equilibrium, they gave robust arguments for the future of chemical kinetics.

After Guldberg and Waage's work, there were other notorious works to lay the fundamental foundations of chemical kinetics, such as the meticulous works of Harcourt and Esson<sup>55</sup>, in order to correlate the concept of chemical equilibrium and temperature dependence with the rate of a chemical reaction. But it was with the proposals drawn up by renowned chemist Jacobus H. van't Hoff<sup>56</sup>, winner of the 1901 Nobel Prize in chemistry *"in recognition of the extraordinary services he has rendered by the discovery of the laws of chemical dynamics and osmotic pressure in solutions."*<sup>57</sup> that chemical kinetics was consolidated. One of van't Hoff's great contributions was the correlation constructed between the enthalpy of the ( $\Delta H$ ) and the displacement of the chemical equilibrium constant ( $K_{eq}$ ) depending on the dependence on temperature, as shown in Equation (2):

$$\frac{d \ln K_{eq}}{d\left(\frac{1}{T}\right)} = \frac{-\Delta H}{R}, \quad (2)$$

where  $\Delta H = H_2 - H_1$ ,  $T$  is the absolute temperature,  $R$  is the universal gas constant and  $C$  is an arbitrary constant.

It is well known that the equilibrium constant is the ratio of the rate constant in the forward ( $k_1$ ) and reverse ( $k_{-1}$ ) direction, so van't Hoff showed that the rate constant should obey a similar set of equations:

$$\frac{d \ln k_1}{dT} = \frac{E_1}{RT^2} + C, \quad (3)$$

and

$$\frac{d \ln k_{-1}}{dT} = \frac{E_2}{RT^2} + C. \quad (4)$$

In the following years, Arrhenius<sup>48,49</sup>, in search of solid arguments for his theory of electrolytic dissociation, spent years in Europe's leading chemistry laboratories, with people such as Ostwald, van't Hoff and Boltzmann, and, based on previous work, deduced one of the most important equations of physical chemistry. However, *"her participation in rate measurements was derived from her desire to gain recognition for her theory of electrolyte dissociation and followed Wilhelm Ostwald production of kinetic evidence supporting her."* The integration of equation (3), that is, the reaction of the forward direction, produces, therefore, the famous Arrhenius equation,

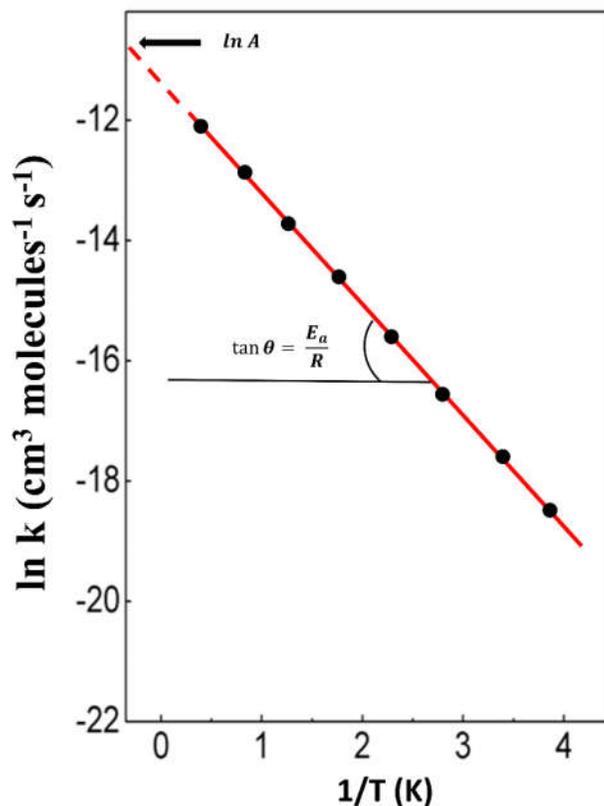
$$k = A e^{\frac{-E_a}{RT}}, \quad (5)$$

where the parameter  $A$  is defined as the frequency constant or pre-exponential factor, and  $E_a$  as the activation energy.

It is important to highlight those other expressions were also proposed, by previously cited authors and by others, to correlate temperature dependence with rate constant. However, it was Arrhenius' expression that had greater agreement with the physical meanings<sup>4,48</sup>. Usually, the Arrhenius equation is written as the natural logarithm of the rate constant being proportional to the reciprocal of temperature:

$$\ln k = \ln A - \frac{E_a}{R} \frac{1}{T}, \quad (6)$$

which in its graphical representation is known as the Arrhenius plot, represented in **Figure 1**.



**Figure 1:** Graphical representation of the Arrhenius plot, where the intercept on the ordinate axis indicates the logarithm of the pre-exponential factor and the tangent of the line provides the ratio of the activation energy by the universal gas constant.

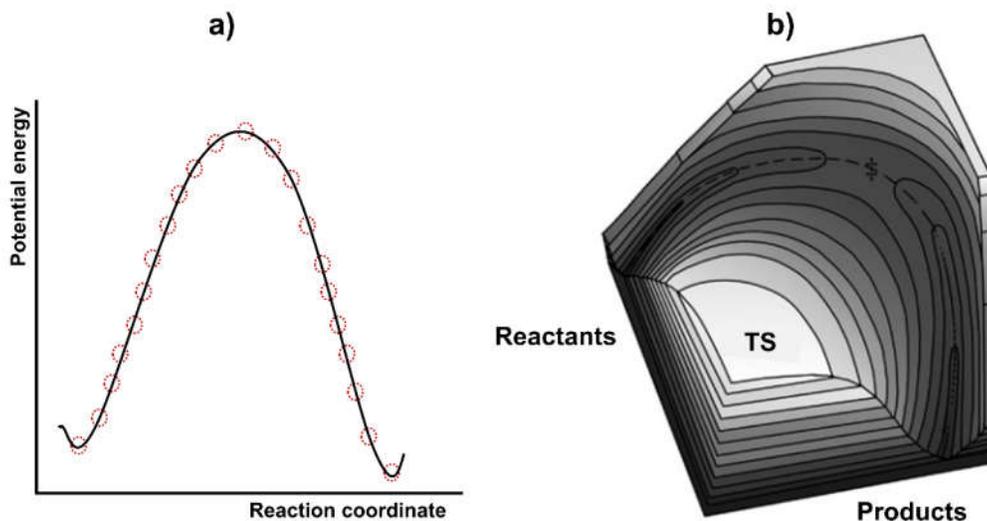
Since Equation (6) is analogous to a linear equation, it is simple to calculate the pre-exponential factor and activation energy of a chemical reaction given experimental data on the rate constant and temperature. Therefore, it is possible to obtain quantitative results from the Arrhenius parameters,  $A$  and  $E_a$ , values which, with their corresponding microscopic properties, did not have a concise explanation until then. The next advances in chemical kinetics were in interpreting the parameters in the Arrhenius law, which was successful but empirical. In order to explain the parameters of Arrhenius, Max Trautz<sup>58</sup> in 1916, and William Lewis<sup>59</sup> in 1918, independently, elaborated the "kinetic theory of gases" based on a recently developed formulation, statistical mechanics, to explain the pre-exponential factor. The hypothesis of this theory considers that there is a collision between the molecules of reagents and that only a fraction of the molecules, with a certain energy, become products<sup>60</sup>. To explain the concept of activation energy, which is crucial in chemical kinetics, some works have been proposed to interpret its meaning<sup>3,4,61</sup>. The work carried out by Marcelin<sup>62</sup>, in 1914, proposed to treat the chemical reaction in terms of movement, but unfortunately did not get continued, due to the early death of this remarkable researcher. Another memorable work was done by Richard C. Tolman<sup>63</sup>, in 1920, which defines the activation energy as the difference between two statistical quantities: the difference between the average energy of all molecular entities with enough energy to react, and the average energy of all molecular entities that reacted or not:

$$\frac{d \ln k}{dT} = \frac{E_{\text{activate}} - E_{\text{average}}}{RT^2} = \frac{E_a}{RT^2} \quad (7)$$

By rearranging the terms, we obtain Equation (8), which is the definition currently recommended by IUPAC <sup>64</sup>:

$$E_a = -R \frac{d \ln k}{d\left(\frac{1}{T}\right)} \quad (8)$$

Despite the success of Tolman's work to interpret the meaning of activation energy, this concept has been confused with other definitions of "energies", such as barrier height, barrier height with zero-point energy correction, internal activation energy, activation enthalpy, among others. Due to the difficulty of correlating and differentiating these concepts, several papers have been published to establish a greater distinction between them and which are highly recommended by the author.<sup>3,61,65</sup> The ideas of Trautz, Lewis and Tolman were essential to the future of chemical kinetics, which resulted in the development of the TST. TST is born concomitantly with the advent of quantum mechanics, which describes the chemical reaction based on the concept of a potential energy surface (PES)<sup>4,8</sup>. The definition of PES results from the Approximation of Born-Oppenheimer which relates to the nuclear movement of the electronic movement due to the disparity of the mass values between the two. To facilitate the interpretation of the concept of PES, see Figure 2a, in which each sphere represents an atomic configuration with its respective coordinate in the phase space and has a value associated with its potential energy. PES can also be interpreted on a topological surface when it comes to more than three variables, like in Figure 2b.



**Figure 2:** Visualization of the potential energy surface for a chemical reaction with a two-dimensional and (b) three-dimensional representation with the main topographic features. Figure 2b has been adapted from reference 1.

## 2.2 Conventional Transition State Theory

Eyring<sup>66</sup>, Evans, Polanyi<sup>67</sup>, Wigner and Pelzer<sup>68,69</sup> independently developed the cornerstone of chemical kinetics, the famous Transition State Theory <sup>70</sup>. TST is a mechanical-

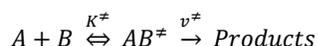
statistical theory, and is the most commonly used theory for calculating velocity constants and interpretations of kinetic data of chemical reactions<sup>9,48</sup>. This statement is simply verified with a simple search with the words "Transition State Theory" in Google Scholar, a most commonly used scientific literature search tool that generates approximately 3.85 million results. These quantitative data confirm the importance of this theory and its relevance.

For the development of TST, Eyring<sup>1</sup> assumed that there was a quasi-equilibrium between the species of the reagents and the species of the transition state and that this balance was maintained along the coordinate of the reaction<sup>4,66</sup>. This new introduced species, the transition state, is defined as a particular configuration that has the highest energy value in one direction and a lower value in a direction perpendicular to the first, known as the saddle point, associated with a potential energy surface. With the hypothesis of quasi-equilibrium a problem with kinetic characteristics, that precisely defines the state of transition, becomes a balancing problem, that were already well-known at the time<sup>48</sup>.

For the derivation of TST, it is important to highlight the three principles on which it was based:

- 1) The reactants and transition state follow a Boltzmann energy distribution law;
- 2) Once species pass through the transition state, they necessarily become products;
- 3) The advance of species on the potential energy surface is considered to be vibrationally and electronically adiabatic and can be handled by classical mechanics; quantum effects are ignored.

For its deduction, the expression of a generic reaction is presented below, which involves the transfer of an atom performed in two steps:



where  $K^\ddagger$  is the equilibrium constant and defined as

$$K^\ddagger = \frac{k_1}{k_{-1}} = \frac{[TS]}{[A][B]}, \quad (9)$$

$\nu^\ddagger$  is the frequency of the species that evolve from the state of transition towards the products due to the movement of a degree of vibrational freedom of the transition state being converted to a degree of translational freedom, and is given by

$$\nu^\ddagger = \frac{k_B T}{h}. \quad (10)$$

The rate law of this reaction can be written as

<sup>1</sup> Usually in the course of the text, the development of the TST will be credited only to Eyring to facilitate writing, which is also generally used in scientific articles and books. However, it is worth mentioning that the development of this formulation had the collaboration of several researchers who have already been cited during the text.

$$v = \frac{d[\text{Products}]}{dt} = v^\ddagger [\text{AB}^\ddagger]; \quad (11)$$

rearranging the terms, you get:

$$v = v^\ddagger K^\ddagger [\text{AB}^\ddagger]. \quad (12)$$

From statistical mechanics, we know that the equilibrium constant can be written according to partition functions, as follows:

$$K^\ddagger = \frac{Q^\ddagger}{Q_A Q_B} \exp(-E_0/RT), \quad (13)$$

where the energy factor counts the zero-point energy difference of the transition state and the reagents at absolute zero, *i.e.*,  $E_0 = E_{\text{AB}^\ddagger} - E_A - E_B$ .

Partition functions provide an indication of the states available at a given temperature in addition to connecting the macroscopic and microscopic properties of a system, and is defined as:

$$Q = \sum w_n \exp\left(-\frac{E_n}{k_B T}\right), \quad (14)$$

where the summation is over all  $n$  states of the system, and the energy can be described with quantum mechanics. The energies and partition functions associated with translational, rotational, vibrational and electronic contributions are shown in Table 1.

Table 1: Contribution types associated with the energy and the partition function.

Type of contribution	Energy	Partition function
Translational	$E_n = \frac{n^2 h^2}{8m x^2}$	$Q_t = \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} V$
Rotational	$E_J = \frac{J(J+1)}{2B} \left(\frac{h}{2\pi}\right)^2$	$Q_r(\text{linear}) = \frac{8\pi^2 I k_B T}{\sigma h^2}$ $Q_r(\text{nonlinear}) = \frac{8\pi^2 (8\pi^3 I_A I_B I_C)^{1/2} (k_B T)^{3/2}}{\sigma h^3}$
Vibrational	$E_v = \left(\frac{1}{2} + n\right) h\nu$	$Q_v = \prod_i \frac{1}{1 - e^{-\frac{h\nu_i}{k_B T}}}$
Electronic	$E_e = E_0$	$Q_e = w_0$

where  $n$  is the quantum number,  $m$  is the mass of the molecule,  $x$  is the barrier width,  $J$  is the rotational quantum number,  $I$  is the moment of inertia,  $\nu$  is the vibrational frequency,  $\sigma$  is the symmetry number and  $w_0$  is the degeneracy of the energy level.

Equation (9) differs from Equation (13) because in the transition state there is the peculiarity that one of its degrees of freedom is connected with the coordinate of the reaction, which allows separation from the other degrees of freedom of the transition state,

$$Q^\ddagger = v^\ddagger Q^\ddagger. \quad (15)$$

Replacing (15) with Equation (12), and knowing that the global velocity law is  $v = k[A][B]$ , provides the famous TST Equation,

$$k_{CTST}(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_A Q_B} \exp(-E_0/RT), \quad (16)$$

### 2.2.1 Transition State Theory in Thermodynamic Formalism

As mentioned earlier, the formulation of conventional transition state theory (CTST) is based on the hypothesis of a quasi-equilibrium between reagents and the transition state, which is a thermodynamic problem. Thus, it is possible to correlate the rate constant through thermodynamic formalism. The deduction of this formalism is usually described using the van't Hoff relationship that correlates Gibbs's free energy with the equilibrium constant as:

$$K^\ddagger = \exp\left(\frac{-\Delta G_0^\ddagger}{RT}\right), \quad (17)$$

where  $\Delta G_0^\ddagger$  is the Gibbs free energy variation for the transition state.

Therefore, we can rewrite the rate constant as follows:

$$k(T) = \frac{k_B T}{h} \exp\left(-\frac{\Delta G_0^\ddagger}{RT}\right). \quad (18)$$

Moreover, it is also useful to write the rate constant as a function of the variation of entropy ( $\Delta S_0^\ddagger$ ) and enthalpy ( $\Delta H_0^\ddagger$ ), the Gibbs free energy can be written according to these thermodynamic properties,  $\Delta G_0^\ddagger = \Delta H_0^\ddagger - T\Delta S_0^\ddagger$ . The rate constant, therefore, is written as:

$$k(T) = \frac{k_B T}{h} \exp\left(\frac{\Delta S_0^\ddagger}{R}\right) \exp\left(\frac{-\Delta H_0^\ddagger}{RT}\right). \quad (19)$$

The formulation of the rate constant correlating thermodynamic properties is a powerful tool to express important information of the transition state from experimental data. For example, the amount of molecules that are located in the transition state can be associated with entropy. And the dependence of enthalpy with the dielectric constant of the solvent, for reactions in a solution, can evidence whether a transition state is neutral, protonated or deprotonated.

Furthermore, the TST written as a function of the Gibbs free energy makes it possible to calculate the generalized transition state, namely, when the transition state is not necessarily located at the saddle point. In this sense, the rate constant is calculated considering that the transition state varies along the reaction coordinate, and when the variational value corresponds to a maximum of  $\Delta G_0^\ddagger$ , the rate constant is minimized.

### 2.3 Corrections to the Transition State Theory

Some corrections were added to the previous equation in order to improve the accuracy of its results with experimental data. One of these corrections is the inclusion of a statistical factor (symmetry factor),  $\sigma$ , which allows concomitance in the PES of several paths in the

reaction. For example, in the H + CH<sub>4</sub> reaction the statistical factor is equal to 4, since the hydrogen radical can abstract any of the four hydrogens from methane. The addition of the symmetry factor in the rate constant considers that this factor is not included in the functions of rotational partitions.

$$k_{CTST}(T) = \sigma \frac{k_B T}{h} \frac{Q^\ddagger}{Q_A Q_B} \exp(-E_0/RT) \quad (20)$$

Some limitations in TST are easily rationalized by analyzing equation (20) and its principles<sup>3,9,48</sup>. It is known that a complete calculation of TST requires some knowledge of PES, since we need to know the value of the barrier height, geometry and vibrational frequencies of both reagents and the transition state. The geometries do not have such a significant effect when compared to the barrier height, since the last term enters the exponential and produces a more significant effect on the rate constant. Vibrational partition functions generally do not lead to discrepancies in calculations, since at common temperatures their value is very close to the unit.

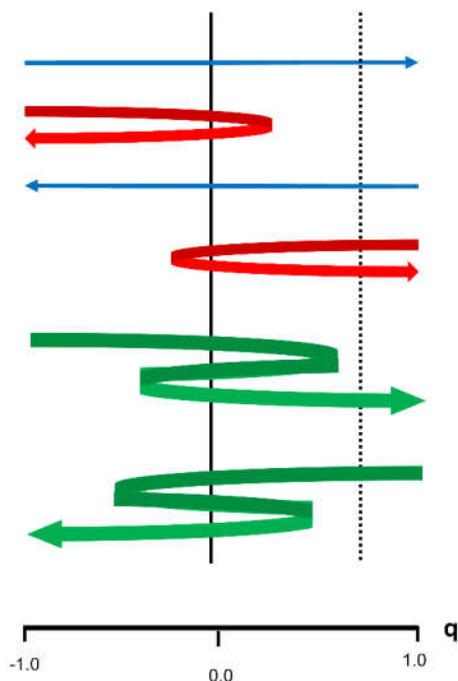
Using the default partition functions as listed in Table 1 can also generate significant errors in rate constant calculations. For example, translational and rotational partition functions are obtained considering that the space between energy levels is very small, compared to the value of  $k_B T$ , which allows the sum to be replaced by the integral, but in some cases this is not valid. Another important factor that can also lead to errors is to consider a harmonic potential in the vibrational partition function, and it is known that at high temperatures there are cases that do not obey a harmonic potential.

The term  $k_B T/h$  expressed in Equation (20) classically accounts for the rate at which reactive systems cross the transition state. However, in real cases, there is a probability that a system will be reflected and the possibility of systems crossing the potential barrier with energy less than that required. These two considerations lead to appreciable errors in the calculation of the rate constant using the CTST formulation and therefore need to be corrected. These corrections are mainly due to their principles, which are corrected when they start to consider variational and quantum effects on the potential energy surface.

### 2.3.1 Variational Transition State Theory

The second hypothesis of conventional TST disregards the possibility of recrossing, that is, once the reagents cross the dividing surface (transition state) they will form the products. However, in a real case, there is a certain probability that the species will return to their initial state. See for example **Figure 3**, which represent trajectories crossing the transition state. In the case of conventional TST, it is considered that there are six reactive trajectories, though in fact only two trajectories are reactive, starting from the reagents (left side) to the products (right side), leading to an overestimated value of TST. The possibility of relocating the transition state, in this case the right, decreases the error in the value of the rate constant. This procedure is called the variational transition state theory (VTST).

The earliest ideals of variational theory were proposed by Keck and Anderson and were extended considerably by Truhlar and his collaborators<sup>9,70-72</sup>. In VTST it is convenient to define a reaction coordinate  $q$  that measures the distance along the minimum energy path. The reagents with  $q = -\infty$  and the products with  $q = +\infty$  and the transition state with a fixed value,  $q = 0$ , are conventional. The variational hypothesis considers that the rate constant is calculated for different transition states ( $q \neq 0$ ), in the case of  $q = 0$  the VTST is reduced to conventional TST<sup>73,74</sup>. Trajectory calculations indicate that recrossing is more significant for collisions involving much more energy than required, usually at high temperatures, or for cases where energy barriers are too low or non-existent<sup>9</sup>.



**Figure 3:** Representative scheme of trajectories that cross the saddle point. The solid line indicates the case of the transition state located in a fixed position, while the dashed line indicates a variational transition state.

VTST can be deduced in two ways<sup>9,70,72</sup>, considering a i) micro-canonical ensemble, which corresponds to calculating a micro-canonical rate constant  $k(E)$  by scanning the transition state in the minimum energy path (MEP), the lowest value is then integrated over all energies, thus providing a value of rate constant; ii) canonical ensemble, the procedure now is to scan not the micro-canonical rate constant but the conventional rate constant, with the lowest value found being accepted as the best estimated value for the rate constant. Garrett and Truhlar<sup>9,70,71</sup> showed that locating the transition state at the point where the Gibbs free energy maximizes activation is the same as minimizing the rate constant, which as a consequence minimizes recrossing, so the equation of rate constant with variational correction can be written as follows:

$$k_{VTST}(T) = \min_q k_{CTST}(T). \quad (21)$$

### 2.3.2 Tunneling corrections

From the point of view of quantum mechanics, there is a certain probability that a system can cross a potential barrier without having the necessary energy, which is not classically permitted and is overlooked in the formulation of CTST<sup>9,10,48</sup>. This effect is called quantum tunneling and can be understood by analyzing the dual behavior of matter (wave-particle) and was interpreted and equated by de Broglie, a physicist, as

$$\lambda = \frac{h}{\sqrt{2mE_c}}, \quad (22)$$

where  $\lambda$  is the wavelength associated with the particle and  $E_c$  is its kinetic energy. To help with the reader's understanding, Table 2 provides an estimate of the wavelength values for different particles associated with their respective masses, which involves the transfer of this particle in a chemical reaction. In this estimate, we assumed a kinetic energy value of 15 kcal mol<sup>-1</sup> (62.76 kJ mol<sup>-1</sup>), which is in accordance with the values available in the literature.

Table 2: Values of wavelengths of different particles with their respective masses, considering a kinetic energy of 15 kcal mol<sup>-1</sup> (62.76 kJ mol<sup>-1</sup>).

Particle	electron	hydrogen	deuterium	carbon	fuleren
mass (kg) <sup>a</sup>	0.0009109	1.6735575	3.3435856	19.944235	1196.6549
$\lambda(\text{Å})$	15.20	0.355	0.251	0.103	0.013

<sup>a</sup> All values are multiplied by 10<sup>27</sup>.

In Table 2, it is unequivocally shown that the electron behaves like a wave, which is also considered significant for the hydrogen atom, making the possibility of tunneling higher. However, when the value associated with the particle mass increases, this undulating behavior and tunneling become negligible. To work around the problem of the tunneling effect that the classic CTST model did not consider, Eyring proposed introducing a multiplicative factor<sup>9,70,75</sup>, called the transmission coefficient  $\kappa(T)$ ,

$$k_{CTST}^T(T) = \kappa(T) \cdot k_{CTST}. \quad (23)$$

To account for the probability of a particle "tunneling" a potential barrier even at lower energies, the superscript  $T$  indicates a tunneling correction factor. In this way, it is necessary to substitute the classical probability,  $P_c(E)$ , of a particle crossing the barrier

$$k_{CTST}(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_A Q_B} \int_0^\infty P_c(E) \exp(-E/k_B T) dE, \quad (24)$$

by a quantum probability,  $P_Q(E)$ , such that

$$k_{CTST}(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_A Q_B} \int_0^\infty P_Q(E) \exp(-E/k_B T) dE. \quad (25)$$

Because of this, the transmission coefficient is commonly defined as the ratio of the equations (24) and (25),

$$\kappa(T) = \frac{\int_0^\infty P_Q(E) \exp\left(-\frac{E_0}{k_B T}\right) d\left(\frac{E}{k_B T}\right)}{\int_0^\infty P_C(E) \exp\left(-\frac{E_0}{k_B T}\right) d\left(\frac{E}{k_B T}\right)} = \int_0^\infty P_Q(E) \exp\left(-\frac{E_0}{k_B T}\right) d\left(\frac{E}{k_B T}\right). \quad (26)$$

The insertion of different models of  $P_Q(E)$  in the previous Equation, in other words, different types of barrier potentials, produces several tunneling corrections, which will be described below.

### 2.3.2.1 Wigner Correction

Eugene Wigner<sup>68</sup>, one of developers of CTST, was also one of the pioneers of tunneling correction. In his approach, he considers a parabolic potential for the movement of nuclei in the region near the top of the barrier, which for lower temperatures, where tunneling is more significant, leads to the following transmission coefficient:

$$\kappa(T) = 1 + \frac{1}{24} \left(\frac{\hbar v^\ddagger}{k_B T}\right)^2, \quad (27)$$

where  $v^\ddagger$  is the imaginary frequency of the transition state, which shows how important it is to have exact imaginary frequency calculations for cases that consider tunneling effects.

### 2.3.2.2 Eckart Correction

Carl Eckart published an article<sup>76</sup> entitled “*THE PENETRATION OF A POTENTIAL BARRIER BY ELECTRONS*” in the renowned journal *The Physical Review* where he presents a new model for calculating the transmission coefficient. He considers a potential barrier, represented in Figure 4, which closely resembles a minimum energy path where there is an exchange of atoms of the type  $A + BC \rightarrow AB + C$ , given by the following expression:

$$V(x) = \frac{Ay}{1-y} - \frac{By}{(1-y)^2}, \quad (28)$$

$$y = \exp\left(\frac{2\pi x}{l}\right), \quad (29)$$

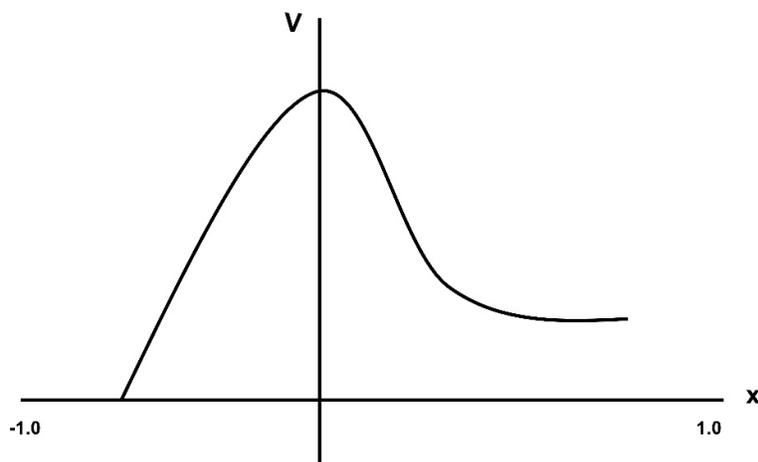
wherein:

$$A = V_{max} - (V_{max} - \Delta V^0), \quad (30)$$

$$B = [V_{max} - \sqrt{V_{max} - \Delta V^0}]^2, \quad (31)$$

$$l = 2\pi \sqrt{\frac{-2}{v^\ddagger} \left[ \frac{1}{\sqrt{V_{max} - \Delta V^0}} + \frac{1}{\sqrt{V_{max}}} \right]^{-1}}, \quad (32)$$

where  $V_{max}$  is the maximum potential of a minimum energy path and  $\Delta V^0 = E_d^0 - E_i^0$ , which is the difference of the barrier height in the direct and reverse direction, respectively.



**Figure 4:** Eckart potential barrier.

Solving the Schrödinger Equation allows the calculation of the probability of a particle with a mass  $\mu$  moving towards an Eckart barrier potential with a given energy  $E$  of  $-\infty$  tunneling through this barrier and later appearing at  $+\infty$  with an energy  $E$ :

$$P_Q(E) = 1 - \frac{\cosh[2\pi(\alpha-\beta)] + \cosh(2\pi\delta)}{\cosh[2\pi(\alpha+\beta)] + \cosh(2\pi\delta)}, \quad (33)$$

where

$$\alpha = \frac{1}{2} \sqrt{\frac{E}{C}}, \quad (34)$$

$$\beta = \frac{1}{2} \sqrt{\frac{E-A}{C}}, \quad (35)$$

$$\delta = \frac{1}{2} \sqrt{\frac{B-C}{C}}, \quad (36)$$

$$C = \frac{\hbar^2}{8\mu l}. \quad (37)$$

In a real situation, it is necessary to count not only a single particle moving toward a potential, but a large amount, with an order of magnitude of  $10^{23}$  particles, with a Boltzmann distribution of energies at a certain temperature, which produces the following expression:

$$\kappa(T) = \exp\left(-\frac{V_{max}}{k_B T}\right) \int_0^\infty P_Q(E) \exp\left(-\frac{E}{k_B T}\right) d\left(\frac{E}{k_B T}\right). \quad (38)$$

### 2.3.2.3 Bell Correction

One of the pioneers in studying the possibility of tunneling in chemical reactions, mainly in hydrogen transfers and its isotopes, was the chemist R. P. Bell<sup>10,77,78</sup>. In 1935, he presented a paper proposing an approximate equation, much simpler than the exact solution of the Schrödinger Equation, to quantitatively analyze the effect of variations on i) the barrier height, ii) the barrier width and iii) particle mass, which, due to the difficulty associated with it, had not yet been addressed.

Two approximations are considered to account for the effects described above. The first is the use of an approximation for the probability of penetration (permeability of the barrier) instead of the exact solution of the Schrodinger Equation, it is used with  $G = 1$  for  $W > E_0$ , and in the case of  $W < E_0$ , the following expression is used:

$$G = \exp\left(\frac{-4\pi\sqrt{2m}}{h}\right) \int_{x_1}^{x_2} [V(x) - W]^{1/2} dx, \quad (39)$$

where  $W$  is the energy of the particle,  $m$  is the mass of the particle,  $V(x)$  is the potential barrier as a function of distance, and  $x_1$  and  $x_2$  are the points of the reaction coordinates for which  $V(x) = 0$ . The second approach is the use of a parabolic potential barrier with a discontinuity at its base,

$$V(x) = E_0 - \frac{E_0 x^2}{a^2}. \quad (40)$$

Replacing Equation (39) in (40), and knowing that  $2a$  is the barrier width, we arrive at:

$$G'_1 = \exp\left(\frac{-2\pi^2 a \sqrt{2m}}{h \sqrt{E_0}} (E_0 - W)\right), \quad (41)$$

$$q = \exp\left(\frac{-E_0}{k_B T}\right) + \frac{1}{k_B T} \int_0^{E_0} \exp\left[\frac{-2\pi^2 a \sqrt{2m}(E_0 - W)}{h \sqrt{E_0}} + \frac{W}{k_B T}\right] dW, \quad (42)$$

$$q = \frac{1}{\beta - \alpha} [\beta \exp(-\alpha) - \alpha \exp(-\beta)], \quad (43)$$

where  $\alpha = \frac{E_0}{k_B T}$  and  $\beta = \frac{2\pi^2 a \sqrt{2m E_0}}{h}$ .

The barrier permeability used above is only valid when  $W$  is significantly less than  $E_0$ , that is, for a high degree of tunneling, but it is imprecise for  $W$  values approximately equal to  $E_0$ . With the aim of providing "a more accurate treatment for a parabolic energy-barrier", Bell, in 1958<sup>78</sup>, proposes a new approach to  $G'_1$  which resembles an exact solution of the Schrodinger Equation using the following conditions:

- i)  $G'_2$  reduces to  $G'_1$  when  $W \ll E_0$  and should tend to drive when  $W \rightarrow \infty$ ,
- ii)  $G(W) = 1/2$  in case  $W = E_0$ , which is a good approximation for realistic cases of interest in chemical kinetics.

The expression that provides the above conditions is represented below:

$$G'_2 = [1 + \exp(\beta y)]^{-1}, \quad (44)$$

With  $y = 1 - W/E_0$ . Replacing the new value of  $G'_2$  in equation (42), we have:

$$q = \int_{-\infty}^1 \frac{\alpha \exp(\alpha y) dy}{1 + \exp(\beta y)}. \quad (45)$$

In the case of  $\alpha < \beta$ , which is the most likely case in situations of chemical interest, and replacing  $x = \exp(\alpha y)$ , Equation (45) can be rewritten as follows:

$$q = \int_0^{e^\alpha} \frac{dx}{1+x^\alpha} = \int_0^\infty \frac{dx}{1+x^\alpha} - \int_{e^\alpha}^\infty \frac{dx}{1+x^\alpha}. \quad (46)$$

Its integration provides:

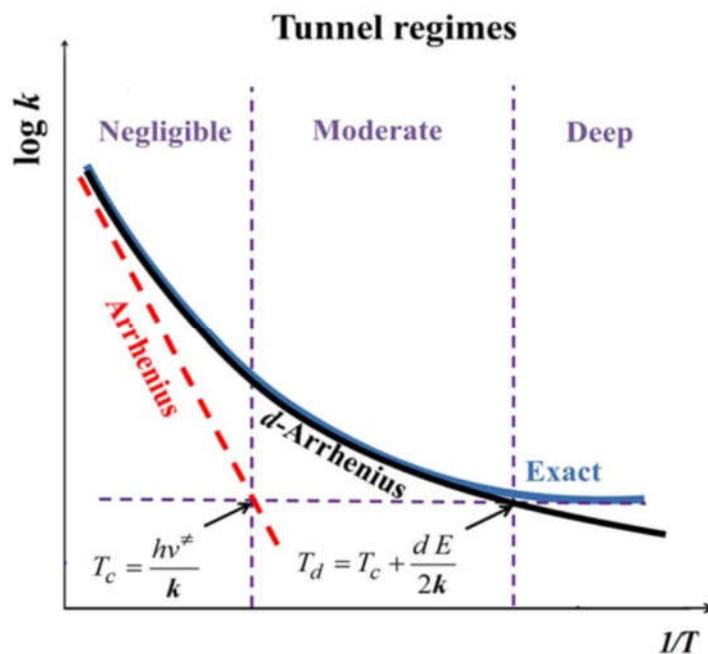
$$q = \frac{\left(\frac{\pi\alpha}{\beta}\right)}{\sin\left(\frac{\pi\alpha}{\beta}\right)} - \frac{\alpha \exp(\alpha - \beta) dy}{\beta - \alpha} \left[ 1 - \frac{\beta - \alpha}{2\beta - \alpha} \exp(-\beta) + \frac{\beta - \alpha}{3\beta - \alpha} \exp(-2\beta) - \dots \right] \quad (47)$$

In most cases,  $\alpha - \beta < 1$ , therefore, the second term of the previous equation can be overlooked, producing:

$$q = \frac{\left(\frac{\pi\alpha}{\beta}\right)}{\sin\left(\frac{\pi\alpha}{\beta}\right)}. \quad (48)$$

#### 2.3.2.4 Deformed Transition State Theory

It is quite consolidated in the literature that a concave curvature in the Arrhenius diagram, as shown in **Figure 5**, is evidence of tunneling in elementary chemical reactions<sup>2,9-11</sup>. In addition, it is known that the tunneling effect on chemical reactions can be classified into four regimes, depending on the value of the crossover temperature<sup>79</sup>,  $T_c = hv^\ddagger/\pi k_B$ . These are: negligible ( $T > 2T_c$ ), small ( $2T_c > T > T_c$ ), moderate ( $T_c > T > T_c/2$ ), and deep ( $T > T_c/2$ ).<sup>78-80</sup>



**Figure 5:** Arrhenius diagram illustrating tunneling regimes: negligible, moderate and deep. The point where the red and purple dotted line intersect is the crossover temperature,  $T_c$ . Figure taken from reference 40.

In order to study chemical processes that had concave curvature in the Arrhenius plot, Aquilanti and Mundim<sup>81,82</sup>, in 2010, proposed an equation to adjust experimental or theoretical data that fit with the available data:

$$k(T) = A \left(1 - d \frac{E_0}{k_B T}\right)^{1/d}, \quad (49)$$

which tends towards the Arrhenius formula when the value  $d$  tends to zero according to the limit proposed by Euler:  $\exp = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$ . However, this proposition is a formula used to provide estimates of values when there is data to be adjusted from macroscopic quantities, such as the rate constant and temperature. In order to obtain a formulation that would allow predictions of the rate constant through phenomenological parameters, Carvalho-Silva and collaborators<sup>2,83</sup>, applied Aquilanti-Mundim formulations to the Transition State Theory, producing the *deformed* Transition State Theory ( $d$ -TST). The rate constant of elementary chemical reactions with sub-Arrhenius behavior is traditionally calculated by introducing a tunneling correction ( $\kappa$ ) in the rate constant of the Transition State Theory (TST) as a multiplicative factor,  $k(T) = \kappa \times k_{TST}$ .

$$k_{d-TST}(T) = \frac{k_B T}{h} \frac{Q^\ddagger}{Q_A Q_B} \left(1 - \frac{dE_0}{k_B T}\right)^{1/d}. \quad (50)$$

Unlike the other formulations, which consider the tunneling parameter as a multiplicative factor (according to Equation (23)), the formulation of the  $d$ -TST replaces the factor  $\kappa \times e^{-E_0\beta}$  formulation of TST by the deformed exponential function. Equation (50) recovers the TST rate constant as  $d$  approaches zero, due to the Euler limit,  $\lim_{d \rightarrow 0} (1 - dE_0\beta)^{1/d} = e^{-E_0\beta}$ , with  $\beta = 1/k_B T$ .

The formulation of  $d$ -TST has clear limitations to describe the deep tunneling regime (Wigner limit) in chemical reactions<sup>2,84</sup>, confirming applicability in negligible to moderate tunneling regimes. For negative and positive values of  $d$  an upward (concave) and downward (convex) curvature, respectively, is observed in the Arrhenius diagram, which are generally called sub- and super Arrhenius behavior. To provide an expression for the deformed parameter ( $d$ ), an analysis of the Bell tunneling correction and the exponential function of the Euler limit were expanded into a power series, which made a connection with the characteristics of the energy barrier possible, yielding

$$d = -\frac{1}{3} \left(\frac{h\nu^\ddagger}{2E_0}\right)^2. \quad (51)$$

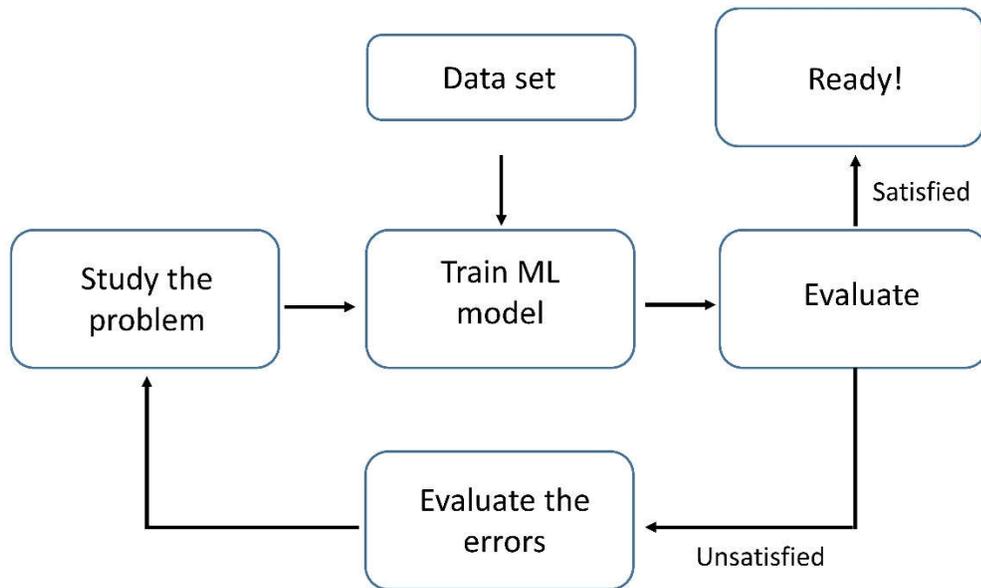
## 2.4 Machine Learning Fundamentals

Turing's question, "Can machines think?" in a seminal paper from the 1950 characterizes the field of Artificial Intelligence (AI)<sup>85</sup>. The construction of techniques to answer what is now called the "Turing Test" provided a great advance in this area. Protocols for building AI models were initially based on two approaches: i) knowledge-based and ii) statistical learning.

Knowledge-based models required cataloged data from experts and code with various conditions was created, however the model was incapable of learning which caused this approach to fall out of use. On the other hand, statistical learning or machine learning uses statistical methods to build models that can learn for themselves, being the main approach used today. In general, machine learning is defined as the area of science aimed at learning from data through computer programming<sup>86</sup>.

However, artificial intelligence and/or machine learning took a few decades to become widely established in use. One of the main bottlenecks in previous decades was the amount of data generated to be used in the models. This challenge has been overcome in recent years with digital transformations driven by industry 4.0, whose main pillars are the internet of things and services, cyberphysical systems and Big-Data.<sup>87,88</sup> In this sense, with the integration of the internet in the main modern services/tasks, the amount of data has grown exponentially. Therefore, the data can be used as a form of raw material, which when processed, can contribute to a set of rules to be taken by the machine algorithms. At this point in the text, it should be noted that even with a significant amount of data, data can commonly be made available with missing information, which makes it essential for the developer to process and pre-process this data to ensure the integrity of the data that will be used. With the data processed, it is essential to carry out an exploratory analysis, in which the main statistical characteristics of the dataset will be summarized. Among these analyses, the following stand out: population and sample, relative and absolute data, frequency distribution, measures of dispersion and position, statistical distributions, among others.

To make it easier for the reader to understand, Figure 6 shows a step-by-step best practice for employing machine learning in various tasks. For example, applied to the case study of this thesis, the first step is to make a detailed study about the problem associated with obtaining the rate constant.



**Figure 6:** An overview for developing a machine learning model.

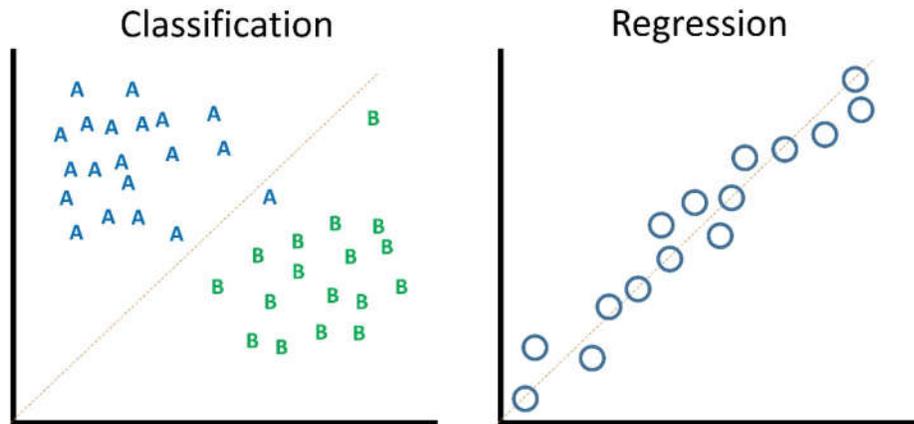
By thoroughly knowing the problem you are studying, patterns will be recognized, as in any operational task. With the recognized patterns and the cataloged dataset – one of the crucial steps for the development of the machine learning model – the model can be trained. The next step is to evaluate whether the model learned correctly, if so, the model can be implemented, otherwise, it is necessary to evaluate the errors and start the studies on the problem again and repeat the procedure described above.

The reader of this thesis could ask at this moment what other tasks, besides the spam case mentioned above, that an artificial intelligence model could perform. To answer this question, it is necessary to divide the types of machine learning into categories, such as:

- i) Supervised Learning: In this category, it is necessary that the training data be provided accompanied with the labels.
- ii) Unsupervised Learning: In this category, training data is not labeled.
- iii) Semi-supervised learning: In this category, the dataset can contain labeled, partially labeled, and unlabeled data.
- iv) Reinforcement learning: In this category, quite different from the other, it is based on teaching tasks to be performed by a learning system through rewards or penalties.

Within supervised learning, the category that will be highlighted in this work, there are two typical tasks: classification and regression. ML models with classification problems are aimed at finding classes, whereas regression problems are aimed at predicting numerical values. Figure 7 shows a didactic example for the afore mentioned cases, a classification problem (on the left) and a regression task (on the right). In the left panel of Figure 7 there would

be a task to predict whether a given letter of the alphabet corresponds to the letter “A” or the letter “B” and in the right panel the task would be destined to perform the prediction, for example, of the value of a house.

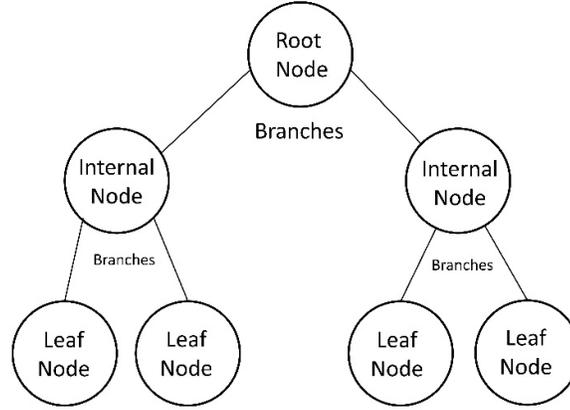


**Figure 7:** Supervised machine learning types.

Purposely, in the case of the classification task, the letter “A” was placed on the side destined for the letter “B”, providing an error to the model. In the case of regression, it is also noticed that some data are far from the dashed line, which represents the model developed. This indicates that the models created need to be evaluated in relation to the mistakes and successes they make. Therefore, statistical metrics to estimate the robustness of models are necessary. In the case of classification problems, the main metrics are precision, accuracy. In the case of regression tasks, the most used metrics are root mean error squared, correlation coefficient.

The knowledge of all these statistical parameters applied to a dataset is defined as good practices for the development of ML models. After representing the data and all necessary treatment, the next step is to apply the ML models. The most important algorithms for supervised learning are: k-nearest neighbors (KNN), linear regression, logistic regression, support vector machines (SVM), decision trees, random forests (RF), XGBoost (XGB) and neural networks (NN)<sup>27</sup>, the last three being the algorithms selected in this thesis. In this sense, a more detailed description of these three algorithms along with the mathematics behind the models is presented below.

Both the “Random Forest” model and the “XGBoost” model are based on algorithms called decision trees. These models are made up of different decision trees, each with its respective nodes, these nodes containing different data that lead to different leaves. The final decision of the model is obtained by an average of all decision trees. Figure 8 illustrates the step by step described above in detail.



**Figure 8:** Architecture of a decision tree-based algorithm.

In the case of regression models, the algorithm uses the mean square error to define how the data will be divided at each node (see Equation 52). On the other hand, in classification cases, the Gini index or entropy can be used – both parameters use probability to determine the result of how the nodes are branched – as shown in Equations 53 and 54.

$$MSE = \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{n}, \quad (52)$$

$$Gini = 1 - \sum_{i=1}^n (p_i)^2, \quad (53)$$

$$Entropy = -\sum_{i=1}^n p_i \cdot \log_2(p_i). \quad (54)$$

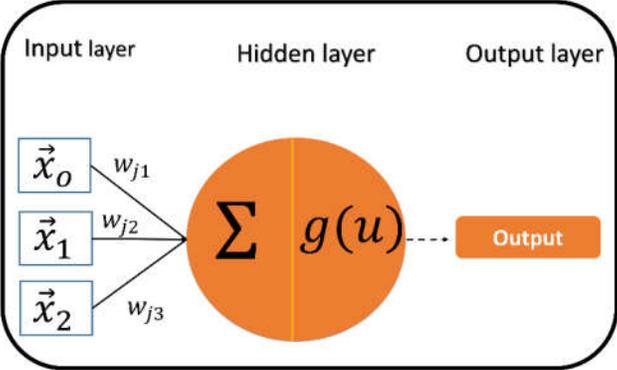
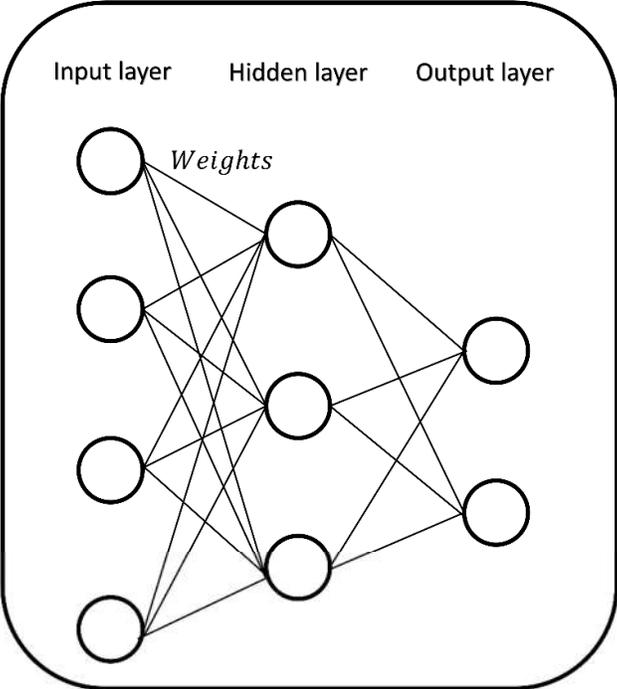
in that  $y_{exp}$ ,  $y_{pred}$ , are the experimental and predicted values and  $p_i$  represents the relative frequency.

Another widely used algorithm for classification and regression problems is neural networks. This algorithm is based on a series of units organized and connected in sequential layers. The neural network architecture involves an input layer, hidden layers and an output layer. The units are the neurons, with neurons within the same layer acting in parallel and transforming the input values received from the previous layer into a scalar value. Figure 9 shows a good illustration for understanding a neural network. The right of Figure 9 represents an overview of the neural network architecture, while the left part shows the example of a neuron being computed separately.

Equation 55 shows the mathematical procedure for calculating each neuron separately.

$$Z = b + \sum_{i=1}^n x_i w_i \quad (55)$$

where  $b$  represents the bias,  $x_i$  the input parameter,  $w_i$  means the weights. Finally, the result of Equation 55 is passed through a nonlinear activation function  $g(z)$ .



**Figure 9:** Processing a neural network.

### 3. PAPERS

The advanced oxidative process is one of the most recommended techniques for the degradation of organic pollutants. Commonly, hydroxyl radical, sulfate radical anion, ozone, singlet oxygen, organic matter, among others, are used in order to degrade these compounds. One of the procedures to verify the degradation efficiency is through the degradation kinetics, since the higher the value of the rate constant, the greater the degradation of the compound. Thus, it is important to obtain this parameter to evaluate the efficiency of the pollutant degradation process, whether in an aqueous environment or in the atmospheric environment. In this thesis, experimental, theoretical and *in silico* approaches were applied to investigate the kinetics, mechanism and toxicity of these compounds through oxidative processes.

The first work was dedicated to evaluating the mechanism, kinetics and toxicity of the herbicide picloram through the oxidative process with the hydroxyl radical. A hybrid approach, that is, theoretical and experimental, was carried out in order to obtain the rate constant and other thermodynamic and kinetic parameters of the degradation reactions. Through this study, it was possible to reveal the main by-products of degradation and evaluate the toxicity of these compounds.

The second and third work were dedicated to the development of machine learning models to predict the rate constant as well as other kinetic parameters to evaluate the degradation of organic pollutants through oxidative processes. In the first of these articles, three machine learning models were developed with the aim of predicting the constant a kinetics of aqueous organic pollutants through the oxidation of hydroxyl radicals and the sulfate radical anion. Similarly, in the second work, the study was extended to the creation of three machine learning models to perform the prediction of atmospheric organic pollutants through the oxidation process initiated by hydroxyl radicals. Additionally, in order to make the developed models available, the two studies were added to a web platform to help professionals and researchers who require information on the degradation efficiency in different environments, especially in water treatment plants.

## PAPER 1:

SANCHES-NETO, F. O., RAMOS, B., LASTRE-ACOSTA, A. M., TEIXEIRA, A. C. S., & CARVALHO-SILVA, V. H. (2021). Aqueous picloram degradation by hydroxyl radicals: **Unveiling mechanism, kinetics, and ecotoxicity through experimental and theoretical approaches**. *Chemosphere*, 278, 130401.

In recent years, several studies have been dedicated to investigating the kinetics of degradation of various compounds in an aqueous environment to assess the consequences of these species in water bodies. Among these compounds, both the use of drugs and pesticides has been of significant concern due to the widespread use of these compounds, either in agriculture or by the general population as a medicine. In the central-west region and in the state of Goiás, the use of herbicides is widely used to control weeds, wheat, barley, oats, and plant species. To combat these pests, one of the most commercially used pesticides is picloram (4-amino-3,5,6-trichloro-2-pyridinecarboxylic acid).

Due to its physicochemical properties, mainly due to its solubility in water, it is essential to evaluate the consequences to the environment caused by the presence of this pesticide in an aquatic environment, since the presence of pesticides has been related to a health problem. public due to its carcinogenic effects on non-target organisms. In this sense, a study revealing the kinetics of degradation through hydroxyl radicals employing a hybrid theoretical-experimental approach was carried out. In addition, details of the main mechanism and toxicity of picloram as well as the main by-products revealed from a theoretical point of view were evaluated using an in-silico methodology.

My contribution in this study was the realization of all the theoretical apparatus for the accomplishment of the electronic structure calculations as well as the kinetic calculations. In addition, all writing is written from the theoretical point of view and the contribution of the results obtained experimentally. The authors Bruno Ramos, Arlen Costa and Antonio Carlos helped to carry out the experimental part. It is also noteworthy that I participated in a technical visit at the University of São Paulo to understand and perform the part experimentally carried out by the authors mentioned above. Professor Valter Carvalho helped to correct and supervise the work carried out.



RightsLink



Home



Help ▾



Live Chat



Sign in



Create Account



## Aqueous picloram degradation by hydroxyl radicals: Unveiling mechanism, kinetics, and ecotoxicity through experimental and theoretical approaches

### Author:

Flávio O. Sanches-Neto, Bruno Ramos, Arlen M. Lastre-Acosta, Antonio Carlos S.C. Teixeira, Valter H. Carvalho-Silva

**Publication:** Chemosphere

**Publisher:** Elsevier

**Date:** September 2021

© 2021 Elsevier Ltd. All rights reserved.

### Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)  
| [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at [customercare@copyright.com](mailto:customercare@copyright.com)



# Aqueous picloram degradation by hydroxyl radicals: Unveiling mechanism, kinetics, and ecotoxicity through experimental and theoretical approaches

Flávio O. Sanches-Neto <sup>a, \*</sup>, Bruno Ramos <sup>b</sup>, Arlen M. Lastre-Acosta <sup>b</sup>,  
Antonio Carlos S.C. Teixeira <sup>b</sup>, Valter H. Carvalho-Silva <sup>a, c, \*\*</sup>

<sup>a</sup> Instituto de Química, Universidade de Brasília, Caixa Postal 4478, 70904-970, Brasília, Brazil

<sup>b</sup> Research Group in Advanced Oxidation Processes (AdOx), Department of Chemical Engineering, Escola Politécnica, University of São Paulo, São Paulo, 05508-010, Brazil

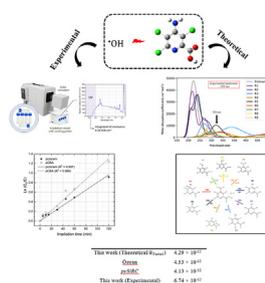
<sup>c</sup> Modeling of Physical and Chemical Transformations Division, Theoretical and Structural Chemistry Group, Research and Postgraduate Center, Goiás State University, 75132-903, Anápolis, Brazil



## HIGHLIGHTS

- Rate constants were estimated by experimental and theoretical approaches.
- Topological analyses and Fukui functions were calculated.
- Toxicity assessment was performed using the ECOSAR computational package.
- The photolysis study was performed by TD-DFT.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 22 January 2021

Received in revised form

9 March 2021

Accepted 22 March 2021

Available online 30 March 2021

Handling Editor: Klaus Kümmerer

### Keywords:

Organic contaminant degradation

Pesticides

d-TST

AOPs

DFT

## ABSTRACT

Pesticides are chemical compounds widely used to combat pests in crops, and they thus play a key role in agricultural production. However, due to their persistence in aquatic environments, even at low concentrations, their use has been considered an environmental problem and caused concern regarding the adverse effects on human health. This paper reports, for the first time, the mechanisms, kinetics, and an evaluation of the toxicity of picloram degradation initiated by  $\cdot\text{OH}$  radicals in the aqueous environment using quantum chemistry and computational toxicology calculations. The rate constants are calculated using a combination of formulations derived from the Transition State Theory in a realistic temperature range (250–310 K). The results indicate that the two favorable pathways (R1 and R5) of  $\cdot\text{OH}$ -based reactions occur by addition to the pyridine ring. The calculated rate constant at 298 K is compared with the overall second-order reaction rate constant, quantified herein experimentally via the competition kinetics method and data available in the literature showing an excellent agreement. The toxicity assessment and a photolysis study provide important information: i) picloram and the majority of degradation products are estimated as harmful; however, ii) these compounds can suffer photolysis in sunlight. The results of the present study can help understand the mechanism of picloram, also providing

\* Corresponding author.

\*\* Corresponding author. Instituto de Química, Universidade de Brasília, Caixa Postal 4478, 70904-970, Brasília, Brazil.

E-mail addresses: [flavio\\_olimpio@outlook.com](mailto:flavio_olimpio@outlook.com) (F.O. Sanches-Neto), [fatiolog@gmail.com](mailto:fatiolog@gmail.com) (V.H. Carvalho-Silva).

important clues regarding risk assessment in aquatic environments as well as novel experimental information.

© 2021 Elsevier Ltd. All rights reserved.

## 1. Introduction

It is well known that the use of pesticides plays a key role in agricultural production (Lin et al., 2018; Tomlin and others, 2009) and in global public health (Organization and others, 1990; Planas et al., 1997), and contributes as an important factor for global economic stability (Seufert et al., 2012; Vasileiadis, 2017). However, in recent years, the widespread use of pesticides has caused significant consequences for the environment and raised concerns due to their carcinogenic and toxic effects on non-target organisms (Canna-Michaelidou and Nicolaou, 1996; Tremolada et al., 2004). Pesticides are complex chemical compounds used to combat pests, such as insects and fungi (Ikehata and El-Din, 2006; Organization and others, 2006). Detailed knowledge of the physical-chemical properties of these species is essential for understanding the impact they might have in the environment (Socorro et al., 2016). The volatility of these compounds, for instance, can contribute to air contamination by evaporation during application (Aktar et al., 2009; Waite et al., 1999); their solubility in water determines the degree to which they can contaminate groundwater (Ghauch, 2001; Luo et al., 2014), and contribute to soil desertification by leaching (Graça et al., 2019) and erosion (Bereswill et al., 2012; Khan, 2016). Many efforts have been devoted to removing pesticides from surface waters, groundwater and industrial effluents because of the adverse effects these species might have on living organisms (An et al., 2014; Cardoso and Valim, 2006). Various technologies, such as dry and wet deposition (Ghauch, 2001; Sauret et al., 2009), adsorption filters (Cardoso and Valim, 2006; Suo et al., 2019), biological treatments (Lafi and Al-Qodah, 2006; Zapata and Oller, 2010), and Advanced Oxidation Processes (AOPs) (An et al., 2014; Oturan and Aaron, 2014) have been developed in a collective effort to remove and/or destroy dangerous contaminants before their disposal in the environment.

Pyridine and derivative compounds have attracted extensive attention due to their occurrence in the environment and the hazardous effects they have on ecosystems and on human health (Abramović et al., 2011; Stapleton et al., 2010). One of the most common pyridine-derived pesticides is picloram (4-amino-3,5,6-trichloro-2-pyridinocarboxylic), a herbicide used for controlling weeds in wheat, barley, oats, and woody plant species (Cardoso and Valim, 2006; Ghauch, 2001; Haag and David Yao, 1992; Hedlund and Youngson, 1972). The toxicity of picloram is considered moderate to high (Abramović et al., 2011; Wauchope et al., 1992), with half-lives in the range 20–300 days (Rahman and Muneer, 2005; Socorro et al., 2016). Furthermore, its photodegradation on the soil surface with the use of aerobic microorganisms is mediated by its efficient solubility in water (Abramović et al., 2011; Ghauch, 2001). All these properties combined with its persistence in the soil confirm the risk of groundwater contamination. In fact, picloram has already been detected in ten American states by the United States Environmental Protection Agency (Ghauch, 2001; Howard, 2017). These features have been supported by several studies of picloram degradation in aqueous solution. Ghauch (2001) studied the degradation of picloram using zero-valent iron powder in an aerobic conical apparatus, in which the pollutant was converted into 4-amino-2-pyridylcarbinol, a substance considered environmentally dangerous. Cardoso and Valim (2006) investigated the

ability of layered double hydroxides to adsorb picloram from aqueous solutions, achieving 96% adsorption after 6 h; however, this physical treatment alone does not promote contaminant degradation. Özcan et al. (2008) used the electro-Fenton process to remove picloram from aqueous solution using current density and catalyst concentration of 300 mA and 0.2 mM Fe<sup>3+</sup>, respectively. The authors reported a reaction rate constant of  $4.53 \times 10^{-12} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$  at 298.15 K. The reaction rate constant of the degradation of picloram with •OH radicals has only been reported at ambient temperatures. Other authors have carried out kinetic and mechanistic studies of picloram photodegradation with titanium dioxide, identifying several intermediates and reaction pathways (Abramović et al., 2011). Recently, Coledam et al. (2018) used four methods based on the production of •OH radicals to evaluate the oxidation and mineralization of picloram: these results support the photo-Fenton HOCl/UVC process as an efficient option to treat aqueous organic contaminants.

The reaction of •OH radicals with organic pollutants is often complex and involves three possible mechanisms (An et al., 2014; Manonmani et al., 2019, 2020; Mei et al., 2019): i) addition of •OH to an aromatic ring or other unsaturated bonds, ii) hydrogen-atom abstraction, and/or iii) single electron transfer. From an experimental perspective, the identification and elucidation of the mechanisms shown above are analytically challenging, complex, expensive and equipment-dependent (An et al., 2014; Mei et al., 2019; Milenković et al., 2020). The laborious experimental procedures involved in distinguishing and quantifying the reaction mechanisms of •OH radicals with organic molecules make quantum theoretical calculations appear as an advantageous protocol to obtain a more detailed picture of the mechanisms and kinetics of such reaction systems. Nevertheless, to the best of our knowledge, there are no theoretical studies regarding the attack of •OH radicals on picloram molecules. Thus, the focus of this work is to provide a detailed understanding of the mechanism and kinetics of picloram degradation mediated by •OH radicals using a blend of quantum chemistry calculations, reaction rate theory, and experimental kinetics procedures. In addition, we provide an evaluation of the photolysis and toxicity of picloram and its degradation products using a TD-DFT procedure and an Ecological Structure-Activity Relationships predictive model.

## 2. Materials and methods

### 2.1. Quantum chemical calculations

The electronic structure properties of the reactants, products, and the transition states were calculated at the M06HF/6-31G+(d) level (for appropriate nomenclature see Fig. 1) with the solvation model density (SMD). The SMD model has been widely used to simulate the aqueous environment in the elucidation of the mechanisms of pesticide degradation, and is computationally less demanding than other continuum models (Luo et al., 2018). Details about other levels of calculations can be found in the Electronic Supplementary Information (ESI) file. The stationary points were characterized by analytic harmonic frequency calculations. The absence or presence of one imaginary frequency characterizes the optimized structures as local minima or transition states,

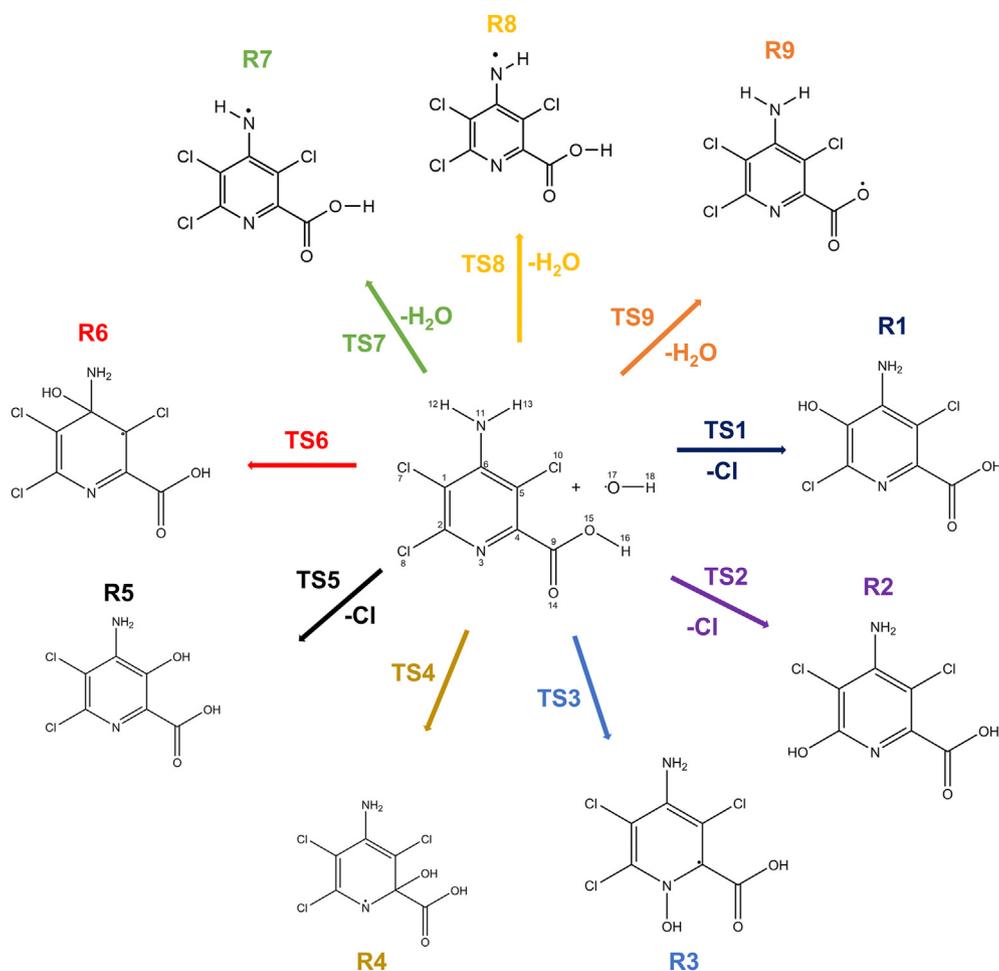


Fig. 1. Scheme of hydrogen atom abstraction and addition reactions with picloram and  $\bullet\text{OH}$  radicals.

respectively. The zero-point vibrational energy contributions have been considered in the calculation of the energy barrier. The photolysis of the optimized geometries was performed by TD-DFT (Gross et al., 1996) calculations using the CAM-B3LYP density functional (Yanai et al., 2004) and 6-311++G(d,p) basis set. Quantum chemical calculations were carried out using the Gaussian 16 package (Frisch et al., 2016).

The topological analysis (Bader, 1985; Matta and Boyd, 2007) was performed in terms of electron density ( $\rho$ ), Laplacian of electron density ( $\nabla^2\rho$ ), Lagrangian kinetic energy density [ $G(r)$ ], Potential energy density [ $V(r)$ ] and Energy density [ $E(r)$ ] at the Critical Points (CP) to efficiently describe H-bonding and its concept without border. To analyze the main reactive sites of the  $\bullet\text{OH} + \text{picloram}$  reaction, appropriate local reactivity descriptors as Fukui function ( $f$ ) (López and Méndez, 2004; Melin et al., 2007)

were calculated according to equations,  $f_{NBO}^- \approx \rho_{NBO}^{LUMO} = \sum_i |c_i|_{LUMO}^2$ ,

and  $f_{NBO}^0 \approx f_{NBO}^+ + f_{NBO}^- / 2$ . The Multiwfn package program (Lu and Chen, 2012) was used to study the topological and Fukui functions.

## 2.2. Reaction rate theory

The reaction rate constant of picloram degradation by  $\bullet\text{OH}$  radicals was calculated using formulations based on the Transition State Theory. To account for the tunneling effect, the deformed Transition State Theory ( $d-TST$ ) (Carvalho-Silva et al., 2017) was

adopted (Eq. (1)):

$$k_{d-TST} = \frac{k_B T}{h} \frac{Q_{TS^i}}{Q_{Reac}} \left(1 - d \frac{E_0}{RT}\right)^{1/d} \quad (1)$$

where  $h$  is the Planck's constant,  $k_B$  is the Boltzmann constant,  $R$  is the universal gas constant,  $d$  is the deformation parameter, while  $Q_{Reac}$  and  $Q_{TS^i}$  are the partition functions of the reactants and transition state, respectively. To include the contribution of molecular diffusion in solution, the calculated rate constant  $k_{d-TST}$  is combined with the steady-state Smoluchowski rate constant,  $k_D$ , following the Collins-Kimball theory (Collins and Kimball, 1949), yielding the apparent rate constant ( $k_{OBS}$ ), according to Eq. (2):

$$\frac{1}{k_{OBS}} = \frac{1}{k_{d-TST}} + \frac{1}{k_D} \quad (2)$$

Additional details about the parameter  $d$  and the calculation of  $k_D$  can be found elsewhere (Sanches-Neto et al., 2020a,b) and the references therein. To consider the effect of recrossing on the reaction rate constant, we also calculate the Variational Transition State Theory (VTST) (Eq. (3)):

$$k_{VTST}(T) = \min_q k_{TST}(T) \quad (3)$$

However, VTST neglects tunneling effects, since it underestimates the kinetic constant for reactions where quantum

tunneling effects are important, especially at low temperatures. Here, in order to account for the quantum tunneling effect, we refine the *deformed* formalism described in [Carvalho-Silva et al. \(2017\)](#), defining the *deformed* Variational Transition State Theory (*d*-VTST) (Eq. (4)):

$$k_{d-VTST}(T) = \min_q k_{d-TST}(T) \quad (4)$$

The Aquilanti-Mundim law ([Aquilanti et al., 2010](#)) (Eq. (5)) was used to fit the rate constant data to represent our results in order to compare with other works:

$$k(T) = A \left( 1 - \bar{d} \frac{\bar{E}}{RT} \right)^{1/\bar{d}} \quad (5)$$

where  $A$  and  $\bar{d}$  are the pre-exponential factor and the deformed parameter, respectively [Note a change in the notation here, needed in order to avoid ambiguities: in terms of the fitted equation, we defined  $\bar{d}$ , which is different from the  $d$  parameter; and  $\bar{E}$ , which is different from  $E_0$ ]. All kinetic and associated parameters have been calculated with the Transitivity Code-version 1.0.4 ([Machado et al., 2019](#)). Details of the computational program can be found on the [www.vhcsgroup.com/transitivity](http://www.vhcsgroup.com/transitivity) web page.

### 2.3. Experimental approach

The overall second-order reaction rate constant between picloram (PCL) and  $\bullet\text{OH}$  radicals ( $k_{PCL,\bullet\text{OH}}$ ) was evaluated using the competition kinetics method with correction for photolysis, as reported elsewhere ([Lastre-Acosta et al., 2019](#); [Shemer et al., 2006](#); [Silva et al., 2015](#); [Son et al., 2020](#); [Wenk et al., 2011](#); [Yan et al., 2021](#)). In this method, the rate constant is evaluated as a function of its observed pseudo-first-order rate in the presence of a competing  $\bullet\text{OH}$  radical scavenger with known kinetics (p-chlorobenzoic acid, pCBA), according to Eqs. (7) and (8):

$$k_{PCL,\bullet\text{OH}} = \left( \frac{k_{PCL}(\text{obs}) - k_{PCL}(\text{dp})}{k_{pCBA}(\text{obs}) - k_{pCBA}(\text{dp})} \right) \times k_{pCBA,\bullet\text{OH}} \quad (6)$$

where  $k_{PCL}(\text{obs})$  is the measured pseudo-first-order reaction rate of picloram in the  $\bullet\text{OH}$  radical system (described below);  $k_{PCL}(\text{dp})$  and  $k_{pCBA}(\text{dp})$  are the measured photolysis rate constants of picloram and p-chlorobenzoic acid.  $k_{pCBA}(\text{obs})$  and  $k_{pCBA,\bullet\text{OH}}$  are, respectively, the measured pseudo-first-order reaction rate constant of the reference compound in the  $\bullet\text{OH}$  radical system and the second-order rate constant of the reaction between this compound and hydroxyl radicals ( $k_{pCBA,\bullet\text{OH}} = 5 \times 10^9 \text{ L} \cdot \text{mol}^{-1} \cdot \text{s}^{-1}$ ) ([Elovitz and Von Gunten, 1999](#)).

The reactional system used in this experiment adopts hydrogen peroxide as a precursor of  $\bullet\text{OH}$  radicals. A reaction mixture is prepared containing  $2.1 \times 10^{-5} \text{ mol} \cdot \text{L}^{-1}$  of picloram (ca. 5 ppm), an equimolar amount of pCBA (ca. 3.2 ppm) and excess ( $0.05 \text{ mol} \cdot \text{L}^{-1}$ ) hydrogen peroxide in natural pH (~4.5).  $\text{H}_2\text{O}_2$  was added in excess to ensure that the competing reactions will be limited by the concentration of the target and the reference compounds. According to the literature, the concentration of hydrogen peroxide was shown not to affect significantly the value of the second-order rate constant of the target compound with  $\bullet\text{OH}$  radicals ([Shemer et al., 2006](#)). The test solution was prepared using deionized, ultrapure water (Milli-Q®, 18.2 MΩ), and distributed into 2.0-mL Pyrex vials with no headspace. The vials were irradiated under simulated sunlight with standard AM1.5G spectra (PEC-L01, Peccell Inc.), as illustrated in [Fig. 2](#), for selected exposure times.

Samples were irradiated for 5, 10, 15, 30, 45, 60 and 120 min. All chemicals, HPLC grade, were acquired from Sigma-Aldrich and used as received without further purification.

The (direct) photolysis rate constants,  $k_{PCL}(\text{dp})$  and  $k_{pCBA}(\text{dp})$ , are evaluated in the same experimental setup, in order to account for the effects of the irradiated photons on the degradation of both species due to excited-state reactions. However, in these analyses, the sample vials are filled with solutions of each compound (5 ppm for picloram and 10 ppm for pCBA) separately and without the addition of  $\text{H}_2\text{O}_2$ . The concentrations of pCBA and picloram were measured using a high-precision liquid chromatography system (LC-10, Shimadzu Co.) equipped with a photodiode array detector (SPD-20MA, Shimadzu Co.). The separation was carried out in a C18 reverse-phase column (Luna C18, 5 μm, 250 × 4.6 mm, Phenomenex Inc.) with isocratic elution of methanol and water (50:50) at 1.5 mL·min<sup>-1</sup>. Picloram and pCBA were detected simultaneously after 4.7 and 13.7 min of elution, respectively, and quantified by UV absorption at 254 nm.

### 2.4. Toxicity assessment

The ecotoxicity of picloram and its main degradation products were determined using the Ecological Structure-Activity Relationship Model (ECOSAR V2.0). ECOSAR is an effective predictive program and has been successfully applied to the ecotoxicity assessment of organic contaminants ([Reuschenbach et al., 2008](#); [Sanderson et al., 2003](#)). Three aquatic organisms – green algae, daphnia, and fish – were chosen as targets. Acute toxicity (feature characterized by LC<sub>50</sub> and EC<sub>50</sub> values) and chronic toxicity (defined by ChV) of the compounds studied were obtained from ECOSAR platform. LC<sub>50</sub> means the concentration of a chemical compound (in mg·L<sup>-1</sup>) that causes the death of half of the fish and daphnia population after exposures of 96 and 48 h, respectively. In addition, EC<sub>50</sub> represents the concentration that permits 50% of green algae to grow normally after 96 h of exposure (in mg·L<sup>-1</sup>).

## 3. Results and discussion

### 3.1. Experimental

The first-order degradation profiles of picloram and pCBA in the hydroxyl radical reaction system are shown in [Fig. 3](#). Both profiles adjusted well to a first-order kinetics, according to Eqs. (7) and (8):

$$\ln \left( \frac{C_0(\text{PCL})}{C(\text{PCL})} \right) = k_{PCL}(\text{obs}) t_{\text{irrad}} \quad (7)$$

$$\ln \left( \frac{C_0(\text{pCBA})}{C(\text{pCBA})} \right) = k_{pCBA}(\text{obs}) t_{\text{irrad}} \quad (8)$$

The reaction rate constants evaluated at this stage are used to calculate  $k_{PCL,\bullet\text{OH}}$ , together with the reaction rates evaluated in the absence of  $\text{H}_2\text{O}_2$ , in order to exclude the effects of direct photolysis, as indicated in Eq. (6). [Table 1](#) summarizes the kinetic constants measured experimentally.

The second-order reaction rate constant is within the range commonly found for the reaction of aromatic compounds and pyridines with hydroxyl radicals ([Buxton et al., 1988](#)), typically between  $8.0 \times 10^{-13}$  and  $2.5 \times 10^{-11} \text{ cm}^3 \cdot \text{molecules}^{-1} \cdot \text{s}^{-1}$ . [Özcan and colleagues \(Özcan et al., 2008\)](#) have carried out a similar competitive kinetics experiment for picloram and reported a second-order rate constant in the same order of magnitude, albeit slightly slower than ours ( $5.64 \times 10^{-12} \text{ cm}^3 \cdot \text{molecules}^{-1} \cdot \text{s}^{-1}$ ). This difference is expected, since our reactive media are substantially different, particularly in terms of pH. As shown in [Buxton's](#)

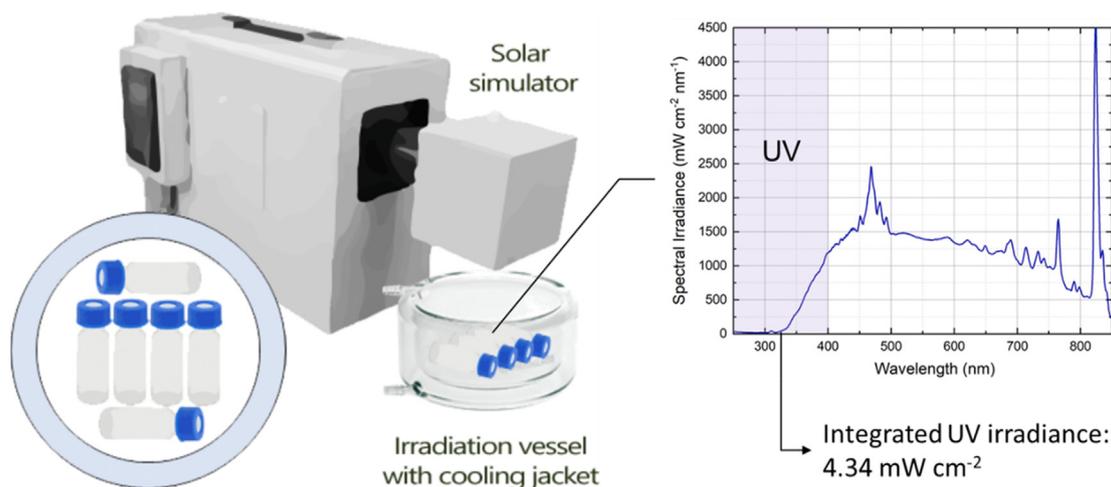


Fig. 2. Illustration of the setup used in the competition kinetics and photolysis experiments.

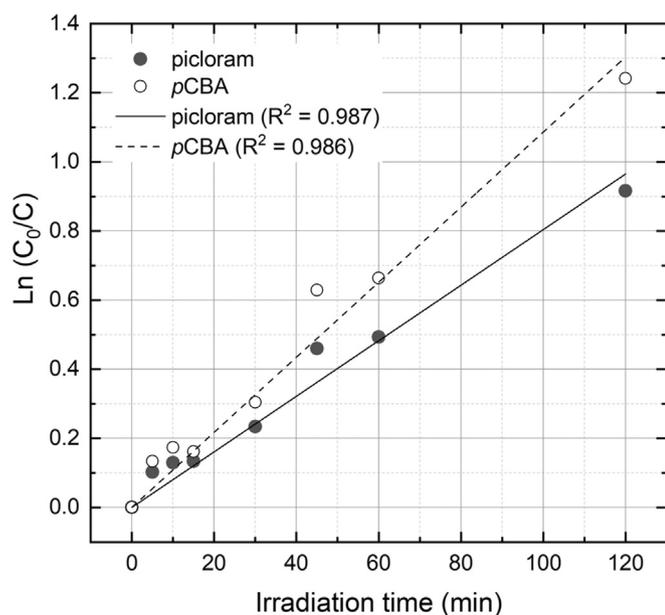


Fig. 3. Pseudo-first-order plots of picloram and pCBA in the hydroxyl radical reaction system.

comprehensive database of kinetic constant rates of organic species with oxygen radicals (Buxton et al., 1988), the measured reaction rates can vary within the same order of magnitude for similar organic compounds according to the reaction pH. Considering the reported pKa of picloram (3.4) (Spadotto and Hornsby, 2003), it is expected that at natural pH most picloram molecules will be in the ionic state; whereas at a pH 3.0, the condition used by Özcan, most of their equivalents would be in the neutral molecular form.

Table 1  
Kinetic parameters obtained experimentally.

Rate constant		Picloram
Direct photolysis	$k_{PCL(dp)}$	$(2.71 \pm 1.95) \times 10^{-6} \text{ s}^{-1}$
•OH (pseudo first-order)	$k_{PCL(obs)}$	$(1.30 \pm 0.04) \times 10^{-4} \text{ s}^{-1}$
Second-order	$k_{PCL, \bullet OH}$	$(6.74 \pm 0.13) \times 10^{-12} \text{ cm}^3 \cdot \text{molecules}^{-1} \cdot \text{s}^{-1}$

### 3.2. Mechanism and energetic parameters

To discuss the mechanisms of the reaction of •OH radicals with picloram, we used the M06HF density functional, which is widely used to study chemical reactions and provides a reliable mechanism and kinetic results (Sanches-Neto et al., 2017, 2020a,b), combined with the 6-31+G(d) base function. Additional calculations with larger basis sets and other DFT functionals were considered, whose results are shown in Table S1. The possible mechanisms for the reaction of •OH radicals with picloram are (see nomenclature in Fig. 1): i) hydrogen transfer from the amino group (R7-R8) or carboxylic group (R9) of picloram by the •OH radical and/or ii) •OH addition to the picloram pyridinic ring (R1-R6). In this study, the single-electron transfer mechanism (SET) is not considered because the barrier height of SET is higher than the reactions involving abstraction and addition (Han et al., 2014; Li et al., 2020; Yang et al., 2017). To confirm the main active sites of the picloram molecule through radical attack, the Fukui functions were calculated, an important approach to explain the reactivity in chemical systems (López and Méndez, 2004; Melin et al., 2007; Milenković et al., 2020; Silva et al., 2010). Fig. 4 illustrates significant values (see nomenclature in Fig. 1) for the selected atoms. According to Fukui formulation (López and Méndez, 2004), the highest values are related to a probable radical attack on carbon C1, C4, and C5 (Fig. 5).

Fig. 3 shows the relative energy profile of the reaction of picloram with •OH radicals calculated at the M06HF/6-31+G(d) level of theory and corroborates the results of the Fukui function – the attack of the •OH radicals on carbons C1, C4, and C5 is kinetically favorable: R1, R4, and R5 channels presented the lowest barrier heights. The Cartesian coordinates of the transition states and picloram calculated in this work are listed in Table S2. Geometric parameters and imaginary frequencies of the transition states involved in the reaction of •OH radicals with picloram are listed in Table S3. From the data in Table S3, it is possible to observe that the transition state geometries for R1 and R5 channels are

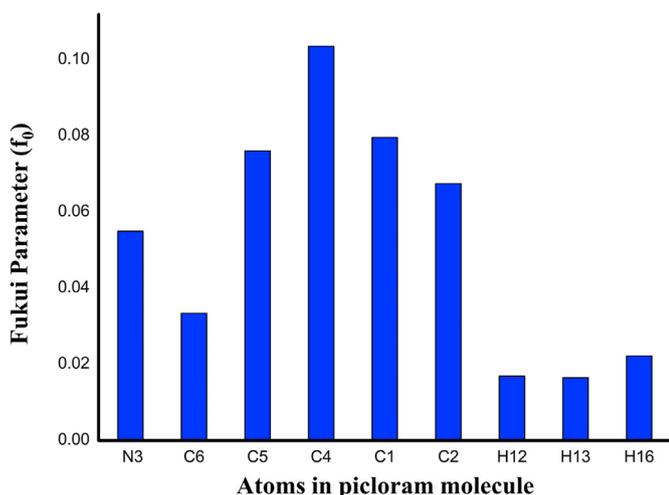


Fig. 4. Fukui function values for the •OH radical attack on selected carbon atoms of picloram in water (see nomenclature in Fig. 1).

similar: at the M06HF/6-31+G(d) level, we found the C<sub>1</sub>–O<sub>17</sub>–H<sub>18</sub> and C<sub>5</sub>–O<sub>17</sub>–H<sub>18</sub> angles of 102.88° and 95.66° and the frequency values for TS1 and TS5 of 650.13i and 681.27i, respectively; this information is very important in modeling the global potential

energy surface for future calculations (Truong et al., 1989). The thermodynamic profile of all the pathways presented in Fig. 1 (R1–R9) was studied and revealed that only the R3 channel is endothermic. The formation of R1, R2, and R5 products occurs by addition of the •OH radical to the picloram ring, forming an intermediate, followed by reductive elimination of the chlorine atom (Abramović et al., 2011; Özcan et al., 2008). The R5 product (4-amino-2,3-dichloro-5-hydroxy-picolinic) has also been reported by previous studies (Abramović et al., 2011; Coledam et al., 2018; Özcan et al., 2008; Rahman and Muneer, 2005).

Additionally, we make use of the Quantum Theory of Atoms in Molecules (QTAIM) to explain the strong thermodynamic stability of R1 and R5 products from their intermolecular interaction. These products are the major degradation compounds found experimentally (see more details in Sec. 3.3). According to Rozas' (Rozas et al., 2000) criteria, intermolecular interactions can be classified as: strong, when  $\nabla^2\rho(r) < 0$ ,  $E(r) < 0$ , and  $|V(r)| > E(r)$ ; ii) weak, when  $\nabla^2\rho(r) > 0$ , and  $E(r) > 0$ ; and iii) moderate, when  $\nabla^2\rho(r) > 0$ , and  $E(r) < 0$ . According to Table 2, there is a strong hydrogen bond between H<sub>18</sub> of the hydroxyl group and O<sub>15</sub> of the carboxylic group permitting an efficient stabilization of the R5 product. Cl<sub>10</sub> in the R1 product provokes a concomitant stabilization mediated by H<sub>15</sub> and O<sub>17</sub> atoms. Molecular representations of critical points (CP) shown in Table 2 are in Fig. S1 of ESI.

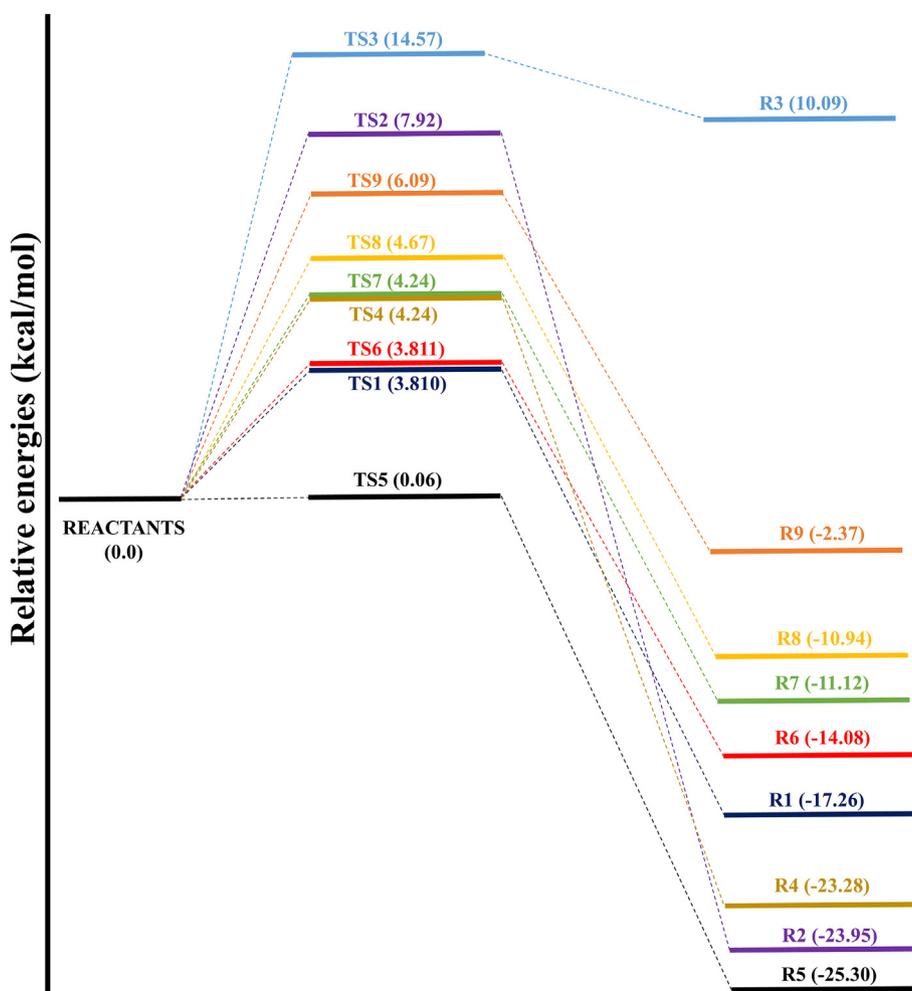


Fig. 5. Relative energy profile corresponding to the initial abstraction of the hydrogen atom and picloram addition reaction by the •OH radical.

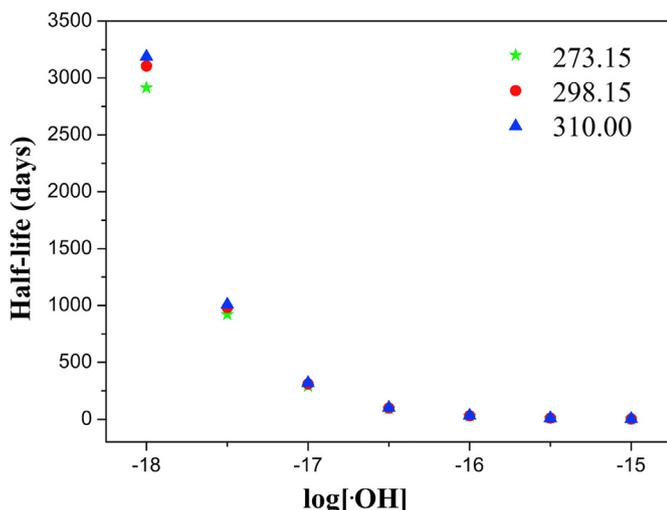


Fig. 6. Half-life time, in days, of picloram degradation as a function of  $\bullet\text{OH}$  concentration, in  $\text{mol}\cdot\text{L}^{-1}$ , in the temperature range 273.15–310 K in natural waters.

### 3.3. Theoretical reaction rate constant

Several authors have suggested that the reductive elimination steps after the addition of the  $\bullet\text{OH}$  radical are not determinant for the calculation of the rate constant (Abramović et al., 2011). Accordingly, reaction rate constants calculated in the present study considered only  $\bullet\text{OH}$  radical additions in the picloram ring and hydrogen abstraction from the amino and carboxylic groups. To the best of our knowledge, these are the first theoretical calculations of the reaction rate constant of picloram degradation mediated by  $\bullet\text{OH}$  radicals.

First, the reaction rate constants of picloram were calculated using the *deformed* Transition State Theory (*d*-TST) in a realistic temperature range (250.0–310.0 K) at the M06HF/6-31+G(d) level. The results are presented in Table 2. The value of the total reaction rate constant at 298.15 K is  $4.29 \times 10^{-12} \text{ cm}^3 \text{ molecules}^{-1} \text{ s}^{-1}$ , which is in excellent agreement with the value found experimentally in this work and by Özcan et al. (2008),  $4.53 \times 10^{-12} \text{ cm}^3 \cdot \text{molecules}^{-1} \cdot \text{s}^{-1}$ . It is observed that the temperature dependence of the reaction rate constants of picloram degradation by  $\bullet\text{OH}$  attack exhibits anti-Arrhenius behavior – a decrease in the reaction rate constant as the temperature increases. Previous studies have shown that reactions with hydroxyl radicals often exhibit deviations from the Arrhenius law. For cases with anti-Arrhenius behavior, the reactional process is characterized by a stereodirectional factor (Coutinho et al., 2015, 2016, 2018). These findings clarify the favoring of the attack of the  $\bullet\text{OH}$  radical to  $\text{C}_5$  of picloram. The substitution of an electron-withdrawing group (Cl) by an electron-donating group ( $\bullet\text{OH}$ ) results in a strong stabilization due to the  $\text{OH}-\text{H}_{18}$  hydrogen bond with the R5 product. These observations support the role of the orientational factor in this reaction.

Recently, we developed a web application structured in a

Table 2

Topological parameters of critical point density, calculated at the theory level M06HF/6-31+G(d) for R1 and R5 products. The parameters  $\rho(r)$ ,  $G(r)$ ,  $V(r)$ ,  $E(r)$  and  $\nabla^2\rho(r)$  are in atomic units.

Interaction	Product	$\rho(r)$	$G(r)$	$V(r)$	$E(r)$	$\nabla^2\rho(r)$
$\text{Cl}_{10}-\text{O}_{17}$	R1	0.01270	0.01037	-0.00155	-0.00882	0.001548
$\text{Cl}_{10}-\text{H}_{13}$	R1	0.01607	0.01442	-0.00304	-0.01137	0.003045
$\text{H}_{18}-\text{O}_{15}$	R5	0.04136	0.03584	0.00273	-0.03856	-0.00273

machine learning and molecular fingerprint algorithm for the estimation of the reaction rate constants of the degradation of organic pollutants in aqueous environments - the pySiRC platform ([www.pysirc.com.br](http://www.pysirc.com.br)) (Sanchez-Neto et al., submitted). Table 3 shows the reaction rate constant estimated by the Bagging machine learning (ML) model in pySiRC. There is an excellent agreement between our quantum chemistry protocol and the ML algorithm.

The branching ratios ( $I_j$ ) were calculated, defined as the ratio of the rate constant of a specific channel and the global rate constant for each channel ( $I_j = k_j/k_{\text{Total}}$ ). The values of the major contributions are given in Table 4, and the results indicate the preference for channels R1 and R5 with branching ratios of 42.9% and 57.1%, at 298.15 K, respectively. The knowledge of the effective importance of each channel is significant and allows the role played by each of these products in the environment to be evaluated.

The reaction rate constant ( $k_{\text{OBS}}$ ) were corrected using the Collins-Kimball formulation to account for diffusion effects. The values of the total diffusion constant ( $k_D$ ) of the reaction are one or two orders of magnitude higher than the reaction rate constants of R1 and R5 channels (see Table S4), which is within the activation-controlled limit; hence, the reaction rate constant is determined by the elementary rate constant. For reactions with energy barrier values close to zero, it is recommended to use the Variational Transition State Theory (VTST) (Bao and Truhlar, 2017; Zhang et al., 2020). Here, we calculated the reaction of the  $\bullet\text{OH}$  radical with picloram for the majority R1 and R5 channels using VTST. For the first time, we also applied the variational correction in the *d*-TST formulation to correct errors due to crossover. The values of the rate constant using the variational correction of TST and *d*-TST, in the range 250.0–310 K are presented in Table S4.

To evaluate the total reaction rate constant of picloram with  $\bullet\text{OH}$  radicals, we adjusted the temperature dependence using the Aquilanti-Mundim law, which has been used successfully to describe the kinetics of chemical processes (Coutinho et al., 2015, 2016) (see Table S5). The profile of equations fitted to Table S5 in the Arrhenius plot showed negative activation energy (anti-Arrhenius kinetics).

From the data of the total reaction rate constant of picloram with the  $\bullet\text{OH}$  radical obtained by *d*-TST, it is possible to calculate the half-life time using  $t_{1/2} = \ln 2 / (k_{\text{total}} \times [\bullet\text{OH}])$ , where  $[\bullet\text{OH}]$  is the concentration of  $\bullet\text{OH}$  radicals in the aqueous media. The half-life of the reaction was studied in the temperature range 273.15–310 K, and  $[\bullet\text{OH}]$  in the range  $10^{-15}$ – $10^{-18} \text{ mol}\cdot\text{L}^{-1}$ , which usually represents the values found in surface waters (Brezonik and Fulkerson-Brekken, 1998; Burns et al., 2012; J. Yang et al., 2020). The calculated half-lives are shown in Fig. 4. From our results, the half-life varies from 31 to 310 days considering a concentration of  $\bullet\text{OH}$  radicals in the range  $10^{-16}$ – $10^{-17} \text{ mol}\cdot\text{L}^{-1}$ , respectively. Our theoretical values are in agreement with the results obtained experimentally, which report a lifetime in the range of 20–300 days (Özcan et al., 2008; Wauchope et al., 1992).

### 3.4. Toxicity evaluation

Many studies that have applied AOPs to remove pesticides

**Table 3**

Reaction rate constants of picloram degradation by  $\bullet\text{OH}$  attack calculated at the M06HF/6-31+G(d) level with the SMD continuous solvation model using *d*-TST formulation. Units in  $\text{cm}^3 \text{molecules}^{-1} \text{s}^{-1}$ .

Rate constant	Temperature (K)				
	250.0	273.15	298.15	300.0	310.0
$k_{R1}$	$2.14 \times 10^{-12}$	$1.98 \times 10^{-12}$	$1.84 \times 10^{-12}$	$1.83 \times 10^{-12}$	$1.79 \times 10^{-12}$
$k_{R2}$	$6.84 \times 10^{-19}$	$2.28 \times 10^{-18}$	$6.84 \times 10^{-18}$	$7.37 \times 10^{-18}$	$1.09 \times 10^{-17}$
$k_{R3}$	$6.18 \times 10^{-25}$	$6.02 \times 10^{-24}$	$4.77 \times 10^{-23}$	$5.49 \times 10^{-23}$	$1.14 \times 10^{-22}$
$k_{R4}$	$8.95 \times 10^{-16}$	$1.53 \times 10^{-15}$	$2.52 \times 10^{-15}$	$2.61 \times 10^{-15}$	$3.12 \times 10^{-15}$
$k_{R5}$	$2.77 \times 10^{-12}$	$2.59 \times 10^{-12}$	$2.45 \times 10^{-12}$	$2.44 \times 10^{-12}$	$2.39 \times 10^{-12}$
$k_{R6}$	$1.04 \times 10^{-17}$	$2.53 \times 10^{-17}$	$5.72 \times 10^{-17}$	$6.05 \times 10^{-17}$	$8.07 \times 10^{-17}$
$k_{R7}$	$3.85 \times 10^{-18}$	$1.01 \times 10^{-17}$	$2.50 \times 10^{-17}$	$2.66 \times 10^{-17}$	$3.69 \times 10^{-17}$
$k_{R8}$	$1.89 \times 10^{-18}$	$4.75 \times 10^{-18}$	$1.130 \times 10^{-17}$	$1.20 \times 10^{-17}$	$1.65 \times 10^{-17}$
$k_{R9}$	$6.52 \times 10^{-22}$	$3.11 \times 10^{-21}$	$1.35 \times 10^{-20}$	$1.49 \times 10^{-20}$	$2.52 \times 10^{-20}$
This work (Theoretical, $k_{Total}$ )	$4.92 \times 10^{-12}$	$4.57 \times 10^{-12}$	$4.29 \times 10^{-12}$	$4.27 \times 10^{-12}$	$4.18 \times 10^{-12}$
Özcan			$4.53 \times 10^{-12}$		
pySiRC			$4.13 \times 10^{-12}$		
This work (Experimental)			$6.74 \times 10^{-12}$		

**Table 4**

Branching ratios, in %, of the elementary channels of picloram degradation using *d*-TST calculated at the M06HF/6-31+G(d) level of theory.

Branching ratio	Temperature (K)				
	250.0	273.15	298.15	300.0	310.0
$\Gamma_{R1}$	43.60	43.30	42.90	42.80	42.80
$\Gamma_{R4}$	0.018	0.034	0.059	0.061	0.075
$\Gamma_{R5}$	56.40	56.70	57.10	57.10	57.10

showed products with toxicity higher than the parent compound (Manonmani et al., 2019, 2020). Computer programs based on an ecological structure-activity relationship model have been widely used to evaluate the environmental and health risks that degradation products might pose. In this work, the toxicity of picloram and the main by-products generated through  $\bullet\text{OH}$  attack were evaluated using the ECOSAR program (Mei et al., 2019; Milenković et al., 2020; J. Yang et al., 2020). Table 5 shows the toxicity classification according to the criteria established by the European Union and China for acute toxicity ( $\text{LC}_{50}$  or  $\text{EC}_{50}$ ) and chronic toxicity (ChV).

The estimated toxicities of the compounds to fish, daphnia, and green algae are reported in Table 6. According to the toxicity parameters, the degradation of picloram via the R1 channel leads to a less toxic product. On the other hand, the R5 product presents higher acute and chronic toxicities than picloram. Recent work evaluating the sulfate-radical oxidation of picloram presented similar toxicity results for picloram and its degradation products (Yang et al., 2020a,b). Interestingly, the R1 and R5 products were not detected when the degradation was carried out with the  $\text{SO}_4^{\bullet-}$  radical. These results reinforce the need to continue the research on the chain of reactions involved in these complex systems to

**Table 5**

Classification of acute and chronic toxicity according to the criteria established by the European Union and Chinese Regulations.

Classification	Acute toxicity <sup>a</sup>	Chronic toxicity <sup>b</sup>
Not harmful	$\text{LC}_{50} > 100$ or $\text{EC}_{50} > 100$	ChV $> 10$
Harmful	$10 < \text{LC}_{50} < 100$ or $10 < \text{EC}_{50} < 100$	$1 < \text{ChV} < 10$
Toxic	$1 < \text{LC}_{50} < 10$ or $1 < \text{EC}_{50} < 10$	$0.1 < \text{ChV} < 1$
Very toxic	$\text{LC}_{50} < 1$ or $\text{EC}_{50} < 1$	ChV $< 0.1$

<sup>a</sup> Criteria set by the European Union (described in Annex VI of Directive 67/548/EEC).

<sup>b</sup> Criteria set by the Chinese hazard evaluation guidelines for new chemical substances (HJ/T 154–2004).

**Table 6**

Toxicity of picloram and its main by-products generated through  $\bullet\text{OH}$  radical attack.

Organisms	Compounds		
	picloram	R1	R5
$\text{LC}_{50}$ (fish 96 h)	682.0	3410	568.0
$\text{LC}_{50}$ (daphnia 48 h)	36.60	65.60	32.40
$\text{EC}_{50}$ (green algae 96 h)	107.0	232.0	93.40
ChV (fish, chronic)	8.920	70.90	7.230
ChV (daphnia, chronic)	0.421	0.680	0.376
ChV (green algae, chronic)	37.90	132.0	32.30

elucidate the main mechanisms of picloram degradation in radical-based AOPs.

### 3.5. Photolysis

The toxicity analysis discussed in the previous section shows that the main by-products can be considered harmful. In order to evaluate an alternative degradation pathway, the photolysis of picloram and its major products was performed using the TD-CAM-B3LYP/6-311++G(d,p) level of calculation, which is widely used to study chemical reactions in the excited state (Kayanuma et al., 2019). Table 7 shows the excitation energy, absorption wavelength, and strength of the harmonic oscillator for picloram and R1 and R5 products.

Vertical excitation energies smaller than 4.13 eV ( $\sim 300$  nm) indicate that the compounds may undergo photolysis under sunlight (Bai et al., 2015) (Fig. 6). From Table 7, both picloram and the R1 product will not photolyze at room temperature. However, the R5 product – the most thermodynamically favorable reaction product but with higher toxicity levels than picloram – may undergo photolysis. Fig. 7 shows the simulated absorption spectra of picloram, with an absorption peak (227 nm) in agreement with experimental observations (223 nm), as well as other degradation products (Dos Santos et al., 2010). A complementary analysis of the

**Table 7**

Vertical excitation energy (eV), absorption wavelength (nm) and oscillator strength (a.u) of picloram and R1 and R5 by-products calculated at the TD-CAM-B3LYP/6-311++G(d,p) level of theory.

Compounds	Excitation energy	Wavelength	Oscillator strength
Picloram	5.4581	227.16	0.8840
R1	5.2080	238.06	0.3774
R5	3.8734	320.09	0.2576

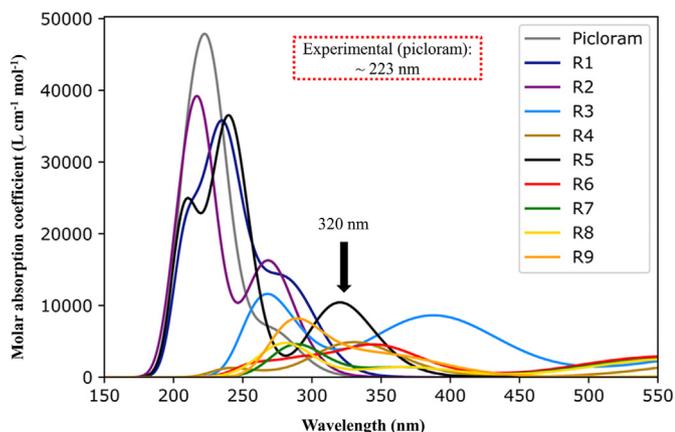


Fig. 7. UV-VIS absorption spectra of picloram and degradation by-products calculated at the (TD)CAM-B3LYP/6–311++G(d,p) level of theory. The experimental value is also presented for comparison (Dos Santos et al., 2010).

bond distances of the R5 product in the ground and first excited state was performed. The values of these optimized parameters are listed in Table S6. Note that the hydrogen bond ( $H_{18}\dots O_{15}$ ) decreases by 0.25 Å, suggesting a possible intramolecular transfer of protons in the excited state. These results show the need for a better understanding of subsequent reactions of hydroxyl radical-based oxidation in aqueous solution.

#### 4. Conclusions

This work presents for the first time a theoretical study of the degradation of picloram by  $\bullet OH$  radicals in aqueous media using a blend of quantum chemistry calculations and reaction rate theories. The calculated values were compared with experimental data obtained by competition kinetics. The mechanisms of picloram degradation and reaction kinetics are reported, and the results show that  $\bullet OH$  addition of picloram occurs favorably at the  $C_1$  and  $C_5$  sites of picloram. The reaction rate constant was calculated with formulations derived from the Transition State Theory. The predicted values for the total rate constant and half-life time at 298.15 K are in very good agreement with experimental results. From the elementary reaction rate constants, the branching ratios of each channel were calculated, accounting R1 and R5 channels as the major by-products. An analysis of the toxicity of picloram and R1 and R5 by-products was performed with the ECOSAR program, showing that these compounds are harmful to living organisms. A photolysis assessment of picloram and R5 by-product indicates that these intermediates can easily be sensitized in sunlight, suggesting additional degradation routes.

#### Author contributions

All authors contributed equally to the planning of the research and the writing of the paper. Flávio O. Sanches-Neto: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Investigation, Data curation; Bruno Ramos: Methodology, Investigation, Data curation; Arlen M. Lastre-Acosta: Methodology, Investigation, Data curation; Antonio Carlos S. C. Teixeira: Writing – review & editing, Supervision, Project administration, Funding acquisition; Valter H. Carvalho-Silva: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors acknowledge the Brazilian funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001), the National Council for Scientific and Technological Development (CNPq, Brazil), and the support of the Sao Paulo Research Foundation (FAPESP Grants 2019/24158–9 and 2018/21721–6) and the Goiás State agency FAPEG for the research funding programs (AUXÍLIO À PESQUISA COLABORATIVA FAPEG-FAPESP/GSP2019011000037). This research is also supported by the High-Performance Computing Center at the Subsecretaria de Tecnologia da Informação (STI), in the Secretaria de Desenvolvimento e Inovação (SEDI), Brazil.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemosphere.2021.130401>.

#### References

- Abramović, B., Šojić, D., Despotović, V., Vione, D., Pazzi, M., Csanádi, J., 2011. A comparative study of the activity of TiO<sub>2</sub> Wackherr and Degussa P25 in the photocatalytic degradation of picloram. *Appl. Catal. B Environ.* 105, 191–198.
- Aktar, Md.W., Sengupta, D., Chowdhury, A., 2009. Impact of pesticides use in agriculture: their benefits and hazards. *Interdiscip. Toxicol.* 2, 1–13. <https://doi.org/10.2478/v10102-009-0001-7>.
- An, T., Gao, Y., Li, G., Kamat, P.V., Peller, J., Joyce, M.V., 2014. Kinetics and mechanism of  $\bullet OH$  mediated degradation of dimethyl phthalate in aqueous solution: experimental and theoretical studies. *Environ. Sci. Technol.* 48, 641–648.
- Aquilanti, V., Mundim, K.C., Elango, M., Kleijn, S., Kasai, T., 2010. Temperature dependence of chemical and biophysical rate processes: phenomenological approach to deviations from Arrhenius law. *Chem. Phys. Lett.* 498, 209–213.
- Bader, R.F.W., 1985. Atoms in molecules. *Acc. Chem. Res.* 18, 9–15.
- Bai, F.-Y., Wang, X., Sun, Y.-Q., Pan, X.-M., 2015. Atmospheric chemistry of alkyl iodides: theoretical studies on the mechanisms and kinetics of CH<sub>3</sub>/C<sub>2</sub>H<sub>5</sub>I+NO<sub>3</sub> reactions. *RSC Adv.* 5, 88087–88095.
- Bao, J.L., Truhlar, D.G., 2017. Variational transition state theory: theoretical framework and recent developments. *Chem. Soc. Rev.* 46, 7548–7596.
- Bereswill, R., Golla, B., Strelake, M., et al., 2012. Entry and toxicity of organic pesticides and copper in vineyard streams: Erosion rills jeopardise the efficiency of riparian buffer strips. *Agric. Ecosyst. Environ.* 145, 81–92.
- Brezonik, P.L., Fulkerson-Brekken, J., 1998. Nitrate-induced photolysis in natural waters: controls on concentrations of hydroxyl radical photo-intermediates by natural scavenging agents. *Environ. Sci. Technol.* 32, 3004–3010.
- Burns, J.M., Cooper, W.J., Ferry, J.L., King, D.W., DiMento, B.P., McNeill, K., Miller, C.J., Miller, W.L., Peake, B.M., Rusak, S.A., others, 2012. Methods for reactive oxygen species (ROS) detection in aqueous environments. *Aquat. Sci.* 74, 683–734.
- Buxton, G.V., Greenstock, C.L., Helman, W.P., Ross, A.B., 1988. Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ( $\bullet OH/\bullet O^-$  in Aqueous Solution). *J. Phys. Chem. Ref. Data* 17, 513–886.
- Canna-Michaelidou, S., Nicolaou, A.-S., 1996. Evaluation of the genotoxicity potential (by Mutatox™ test) of ten pesticides found as water pollutants in Cyprus. *Sci. Total Environ.* 193, 27–35.
- Cardoso, L.P., Valim, J.B., 2006. Study of acids herbicides removal by calcined Mg-Al-CO<sub>3</sub>-LDH. *J. Phys. Chem. Solid.* 67, 987–993.
- Carvalho-Silva, V.H., Aquilanti, V., de Oliveira, H.C.B., Mundim, K.C., 2017. Deformed transition-state theory: deviation from Arrhenius behavior and application to bimolecular hydrogen transfer reaction rates in the tunneling regime. *J. Comput. Chem.* 38, 178–188.
- Coledam, D.A.C., Sánchez-Montes, I., Silva, B.F., Aquino, J.M., 2018. On the performance of HOCl/Fe<sup>2+</sup>, HOCl/Fe<sup>2+</sup>/UVA, and HOCl/UVC processes using in situ electrogenerated active chlorine to mineralize the herbicide picloram. *Appl. Catal. B Environ.* 227, 170–177.
- Collins, F.C., Kimball, G.E., 1949. Diffusion-controlled reactions in liquid solutions. *Ind. Eng. Chem.* 41, 2551–2553.
- Coutinho, N.D., Aquilanti, V., Silva, V.H.C., Camargo, A.J., Mundim, K.C., De Oliveira, H.C.B., 2016. Stereodirectional origin of anti-arrhenius kinetics for a tetraatomic hydrogen exchange reaction: born-oppenheimer molecular dynamics for OH + HBr. *J. Phys. Chem.* 120, 5408–5417.

- Coutinho, N.D., Sanches-Neto, F.O., Carvalho-Silva, V.H., de Oliveira, H.C.B., Ribeiro, L.A., Aquilanti, V., 2018. Kinetics of the  $\text{OH} + \text{HCl} \rightarrow \text{H}_2\text{O} + \text{Cl}$  reaction: rate determining roles of stereodynamics and roaming and of quantum tunneling. *J. Comput. Chem.* 39, 2508–2516.
- Coutinho, N.D., Silva, V.H.C., De Oliveira, H.C.B., Camargo, A.J., Mundim, K.C., Aquilanti, V., 2015. Stereodynamical origin of anti-arrhenius kinetics: negative activation energy and roaming for a four-atom reaction. *J. Phys. Chem. Lett.* 6, 1553–1558.
- Dos Santos, L.B.O., Infante, C.M.C., Masini, J.C., 2010. Determination of picloram in waters by sequential injection chromatography with UV detection. *J. Braz. Chem. Soc.* 21, 1557–1562.
- Elovitz, M.S., Von Gunten, U., 1999. Hydroxyl radical/ozone ratios during ozonation processes. I. The R(CT) concept. *Ozone Sci. Eng.* 21, 239–260.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J., 2016. Gaussian16 Revision C.01.
- Ghauch, A., 2001. Degradation of benomyl, picloram, and dicamba in a conical apparatus by zero-valent iron powder. *Chemosphere* 43, 1109–1117.
- Gross, E.K.U., Dobson, J.F., Petersilka, M., 1996. Density functional theory of time-dependent phenomena. *Density Functional Theory II*. Springer, pp. 81–172.
- Haag, W.R., David Yao, C.C., 1992. Rate constants for reaction of hydroxyl radicals with several drinking water contaminants. *Environ. Sci. Technol.* 26, 1005–1013.
- Han, D., Li, J., Cao, H., He, M., Hu, J., Yao, S., 2014. Theoretical investigation on the mechanisms and kinetics of OH-initiated photooxidation of dimethyl phthalate (DMP) in atmosphere. *Chemosphere* 95, 50–57.
- Hedlund, R.T., Youngson, C.R., 1972. The rates of photodecomposition of picloram in aqueous systems. *Adv. Chem. Ser. no. 111*, 159–172.
- Howard, P., 2017. *Handbook of Environmental Fate and Exposure Data: for Organic Chemicals, Volume III Pesticides*. Routledge.
- Ikehata, K., El-Din, M.G., 2006. Aqueous pesticide degradation by hydrogen peroxide/ultraviolet irradiation and Fenton-type advanced oxidation processes: a review. *J. Environ. Eng. Sci.* 5, 81–135.
- Kayanuma, M., Shoji, M., Furuya, K., Aikawa, Y., Umemura, M., Shigeta, Y., 2019. Theoretical study of the photodissociation reaction of methanol. *Chem. Phys. Lett.* 714, 137–142.
- Khan, S.U., 2016. *Pesticides in the Soil Environment*. Elsevier, Amsterdam.
- Lafi, W.K., Al-Qodah, Z., 2006. Combined advanced oxidation and biological treatment processes for the removal of pesticides from aqueous solutions. *J. Hazard Mater.* 137, 489–497. <https://doi.org/10.1016/j.jhazmat.2006.02.027>.
- Lastre-Acosta, A.M., Barberato, B., Parizi, M.P.S., Teixeira, A.C.S.C., 2019. Direct and indirect photolysis of the antibiotic enoxacin: kinetics of oxidation by reactive photo-induced species and simulations. *Environ. Sci. Pollut. Res.* 26, 4337–4347.
- Li, H., Miao, X., Zhang, J., Du, J., Xu, S., Tang, J., Zhang, Y., 2020. DFT studies on the reaction mechanism and kinetics of dibutyl phthalate initiated by hydroxyl and sulfate radicals: prediction of the most reactive sites. *Chem. Eng. J.* 381, 122680.
- Lin, C., Zhang, L., Zhang, H., Wang, Q., Zhu, J., Wang, J., Qian, M., 2018. Enantioselective degradation of myclobutanil and famoxadone in grape. *Environ. Sci. Pollut. Res.* 25, 2718–2725.
- López, P., Méndez, F., 2004. Fukui function as a descriptor of the imidazolium protonated cation resonance hybrid structure. *Org. Lett.* 6, 1781–1783.
- Lu, T., Chen, F., 2012. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592.
- Luo, S., Gao, L., Wei, Z., Spinney, R., Dionysiou, D.D., Hu, W.P., Chai, L., Xiao, R., 2018. Kinetic and mechanistic aspects of hydroxyl radical-mediated degradation of naproxen and reaction intermediates. *Water Res.* 137, 233–241.
- Machado, H.G., Sanches-Neto, F.O., Coutinho, N.D., Mundim, K.C., Palazzetti, F., Carvalho-Silva, V.H., 2019. “Transitivity”: a Code for computing kinetic and related parameters in chemical Transformations and transport phenomena. *Molecules* 24, 3478.
- Manonmani, G., Sandhiya, L., Senthilkumar, K., 2020. Mechanism and kinetics of diuron oxidation by hydroxyl radical addition reaction. *Environ. Sci. Pollut. Res.* 27, 12080–12095.
- Manonmani, G., Sandhiya, L., Senthilkumar, K., 2019. Mechanism and kinetics of diuron oxidation initiated by hydroxyl radical: hydrogen and chlorine atom abstraction reactions. *J. Phys. Chem.* 123, 8954–8967.
- Matta, C.F., Boyd, R.J., 2007. *The Quantum Theory of Atoms in Molecules: from Solid State to DNA and Drug Design*. John Wiley & Sons.
- Mei, Q., Sun, J., Han, D., Wei, B., An, Z., Wang, X., Xie, J., Zhan, J., He, M., 2019. Sulfate and hydroxyl radicals-initiated degradation reaction on phenolic contaminants in the aqueous phase: mechanisms, kinetics and toxicity assessment. *Chem. Eng. J.* 373, 668–676.
- Melin, J., Ayers, P.W., Ortiz, J.V., 2007. Removing electrons can increase the electron density: a computational study of negative Fukui functions. *J. Phys. Chem.* 111, 10017–10019.
- Milenković, D.A., Dimić, D.S., Avdović, E.H., Amić, A.D., Dimitrić Marković, J.M., Marković, Z.S., 2020. Advanced oxidation process of coumarins by hydroxyl radical: towards the new mechanism leading to less toxic products. *Chem. Eng. J.* 395, 124971.
- Organization, W.H., others, 2006. *Pesticides and Their Application: for the Control of Vectors and Pests of Public Health Importance*.
- Organization, W.H., others, 1990. *Public Health Impact of Pesticides Used in Agriculture*. World Health Organization.
- Özcan, A., Şahin, Y., Kopal, A.S., Oturan, M.A., 2008. Degradation of picloram by the electro-Fenton process. *J. Hazard Mater.* 153, 718–727.
- Planas, C., Caixach, J., Santos, F.J., Rivera, J., 1997. Occurrence of pesticides in Spanish surface waters. Analysis by high resolution gas chromatography coupled to mass spectrometry. *Chemosphere* 34, 2393–2406.
- Rahman, M.A., Muneer, M., 2005. Heterogeneous photocatalytic degradation of picloram, dicamba, and floumeturon in aqueous suspensions of titanium dioxide. *J. Environ. Sci. Health Part B Pestic. Food Contam. Agric. Wastes* 40, 247–267.
- Reuschenbach, P., Silvani, M., Dammann, M., Warnecke, D., Knacker, T., 2008. ECOSAR model performance with a large test set of industrial chemicals. *Chemosphere* 71, 1986–1995.
- Rozas, I., Alkorta, I., Elguero, J., 2000. Behavior of ylides containing N, O, and C atoms as hydrogen bond acceptors. *J. Am. Chem. Soc.* 122, 11154–11161.
- Sanches-Neto, F.O., Coutinho, N.D., Carvalho-Silva, V.H., 2017. A novel assessment of the role of the methyl radical and water formation channel in the  $\text{CH}_3\text{OH} + \text{H}$  reaction. *Phys. Chem. Chem. Phys.* 19, 24467–24477.
- Sanches-Neto, F.O., Coutinho, N.D., Palazzetti, F., Carvalho-Silva, V.H., 2020a. Temperature dependence of rate constants for the  $\text{H(D)} + \text{CH}_4$  reaction in gas and aqueous phase: deformed Transition-State Theory study including quantum tunneling and diffusion effects. *Struct. Chem.* 31.
- Sanches-Neto, Flávio, O., Coutinho, N.D., Palazzetti, F., Carvalho-Silva, V.H., 2020b. Temperature dependence of rate constants for the  $\text{H(D)} + \text{CH}_4$  reaction in gas and aqueous phase: deformed Transition-State Theory study including quantum tunneling and diffusion effects. *Struct. Chem.* 31, 609–617.
- Sanches-Neto, F.O., S, J.R.D., Junior, L.H.K.Q., Carvalho-Silva, V.H., n.d. “pySiRC”: Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants. submitted.
- Sanderson, H., Johnson, D.J., Wilson, C.J., Brain, R.A., Solomon, K.R., 2003. Probabilistic hazard assessment of environmentally occurring pharmaceuticals toxicity to fish, daphnids and algae by ECOSAR screening. *Toxicol. Lett.* 144, 383–395.
- Sauret, N., Wortham, H., Strekowski, R., et al., 2009. Comparison of annual dry and wet deposition fluxes of selected pesticides in Strasbourg, France. *Environ. Pollut.* 157, 303–312.
- Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. *Nature* 485, 229–232.
- Shemer, H., Sharpless, C.M., Elovitz, M.S., Linden, K.G., 2006. Relative rate constants of contaminant candidate list pesticides with hydroxyl radicals. *Environ. Sci. Technol.* 40, 4460–4466.
- Silva, M.P., Mostafa, S., McKay, G., Rosario-Ortiz, F.L., Teixeira, A.C.S.C., 2015. Photochemical fate of amicarbazone in aqueous media: laboratory measurement and simulations. *Environ. Eng. Sci.* 32, 730–740.
- Silva, V.H.C., Camargo, L., Napolitano, H.B., Pérez, C.N., Camargo, A.J., 2010. Theoretical investigation of the interaction of glycerol with aluminum and magnesium phthalocyanines. *J. Mol. Graph. Model.* 29, 206–213.
- Socorro, J., Durand, A., Temime-Roussel, B., Gligorovski, S., Wortham, H., Quivet, E., 2016. The persistence of pesticides in atmospheric particulate phase: an emerging air quality issue. *Sci. Rep.* 6, 1–7.
- Son, Y., Lee, Y.M., Zoh, K.D., 2020. Kinetics and degradation mechanism of tris (1-chloro-2-propyl) phosphate in the UV/H<sub>2</sub>O<sub>2</sub> reaction. *Chemosphere* 260, 127461.
- Spadotto, C.A., Hornsby, A.G., 2003. Soil sorption of acidic pesticides: modeling pH effects. *J. Environ. Qual.* 32, 949–956.
- Stapleton, D.R., Konstantinou, I.K., Mantzavinos, D., Hela, D., Papadaki, M., 2010. On the kinetics and mechanisms of photolytic/TiO<sub>2</sub>-photocatalytic degradation of substituted pyridines in aqueous solutions. *Appl. Catal. B Environ.* 95, 100–109.
- Suo, F., You, X., Ma, Y., et al., 2019. Rapid removal of triazine pesticides by P doped biochar and the adsorption mechanism. *Chemosphere* 235, 918–925.
- Tomlin, C.D.S., others, 2009. *The Pesticide Manual: A World Compendium*. British Crop Production Council.
- Tremolada, P., Finizio, A., Villa, S., Gaggi, C., Vighi, M., 2004. Quantitative inter-specific chemical activity relationships of pesticides in the aquatic environment. *Aquat. Toxicol.* 67, 87–103.
- Truong, T.N., Truhlar, D.G., Baldrige, K.K., Gordon, M.S., Steckler, R., 1989. Transition state structure, barrier height, and vibrational frequencies for the reaction  $\text{Cl} + \text{CH}_4 \rightarrow \text{CH}_3 + \text{HCl}$ . *J. Chem. Phys.* 90, 7137–7142.
- Vasileiadis, V.P., 2017. Economic sustainability: less pesticide rarely causes loss. *Native Plants* 3.
- Waite, D.T., Cessna, A.J., Gurprasad, N.P., et al., 1999. A new sampler for collecting separate dry and wet atmospheric depositions of trace organic chemicals. *Atmos. Environ.* 33, 1513–1523.
- Wauchope, R.D., Buttler, T.M., Hornsby, A.G., Augustijn-Beckers, P.W.M., Burt, J.P., 1992. *The SCS/ARS/CES pesticide properties database for environmental decision-making. Reviews of Environmental Contamination and Toxicology*. Springer, pp. 1–155.
- Wenk, J., Gunten, U. Von, Canonica, S., 2011. Effect of dissolved organic matter on

- the transformation of contaminants induced by excited triplet states and the hydroxyl radical. *Environ. Sci. Technol.* 45, 1334–1340.
- Yan, B., Xu, D.C., Liu, Z., Tang, J., Huang, R., Zhang, M., Cui, F., Shi, W., Hu, C., 2021. Degradation of neurotoxin  $\beta$ -N-methylamino-L-alanine by UV254 activated persulfate: kinetic model and reaction pathways. *Chem. Eng. J.* 404, 127041.
- Yanai, T., Tew, D.P., Handy, N.C., 2004. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 393, 51–57.
- Yang, J., Wang, Z., Lv, G., Liu, W., Wang, Y., Sun, X., Gao, J., 2020a. Indirect photodegradation of fludioxonil by hydroxyl radical and singlet oxygen in aquatic environment: mechanism, photoproducts formation and eco-toxicity assessment. *Ecotoxicol. Environ. Saf.* 197, 110644.
- Yang, X., Cao, X., Zhang, L., Wu, Y., Zhou, L., Xiu, G., Ferronato, C., Chovelon, J.-M., 2020b. Sulfate radical-based oxidation of the aminopyralid and picloram herbicides: the role of amino group on pyridine ring. *J. Hazard Mater.* 124181.
- Yang, Z., Su, R., Luo, S., Spinney, R., Cai, M., Xiao, R., Wei, Z., 2017. Comparison of the reactivity of ibuprofen with sulfate and hydroxyl radicals: an experimental and theoretical study. *Sci. Total Environ.* 590–591, 751–760.
- Zapata, A., Oller, L., et al., 2010. Decontamination of industrial wastewater containing pesticides by combining large-scale homogeneous solar photocatalysis and biological treatment. *Chem. Eng. J.* 160, 447–456. <https://doi.org/10.1016/j.cej.2010.03.042>.
- Zhang, L., Truhlar, D.G., Sun, S., 2020. Association of Cl with C<sub>2</sub>H<sub>2</sub> by unified variable-reaction-coordinate and reaction-path variational transition-state theory. *Proc. Natl. Acad. Sci. U.S.A.* 117, 5610–5616.

## PAPER 2:

SANCHES-NETO, F. O., DIAS-SILVA, J. R., KENG QUEIROZ JUNIOR, L. H., & CARVALHO-SILVA, V. H. (2021). **“pySiRC”: Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants.** *Environmental Science & Technology*, 55(18), 12437-12448.

The presence of organic pollutants in an aqueous environment has caused great concern to the international scientific community due to adverse effects related to global health. In this sense, the efficient degradation of these pollutants has been the basis of several studies in the literature. One of the widely employed approaches is the use of advanced oxidative processes to degrade these compounds. However, obtaining kinetic parameters, such as the rate constant and the half-life, requires an experimental and theoretical approach. However, as shown by previous work in this thesis, obtaining the rate constant from an experimental point of view is analytically challenging and from a theoretical point of view requires Herculean protocols of chemical kinetics and quantum chemistry.

Therefore, motivated by the difficulty in obtaining the values of the rate constant through theoretical and experimental approaches such as those carried out in this thesis described in the previous section, we developed a study to predict the rate constant using machine learning methods. In addition, we have developed a free and easily accessible web application for the broad access of the scientific community and technical managers to obtain the rate constant with a few clicks.

My contribution in this work was the development of machine learning models and the writing of the article. Author Jefferson Richard contributed richly to the creation of the models and discussions. Professor Luiz Keng contributed to the statistical analyses. Professor Valter Carvalho helped to correct and supervise the work carried out.



RightsLink



Home



Help ▾



Live Chat



Sign in



Create Account

## "pySiRC": Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants

**Author:**

Flávio Olimpio Sanches-Neto, Jefferson Richard Dias-Silva, Luiz Henrique Keng Queiroz Junior, et al

**Publication:** Environmental Science & Technology

**Publisher:** American Chemical Society

**Date:** Sep 1, 2021

*Copyright © 2021, American Chemical Society*

### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)  
| [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at  
[customer care@copyright.com](mailto:customer care@copyright.com)

# “pySiRC”: Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants

Flávio Olímpio Sanches-Neto,\* Jefferson Richard Dias-Silva,\* Luiz Henrique Keng Queiroz Junior,\* and Valter Henrique Carvalho-Silva\*



Cite This: *Environ. Sci. Technol.* 2021, 55, 12437–12448



Read Online

ACCESS |



Metrics & More



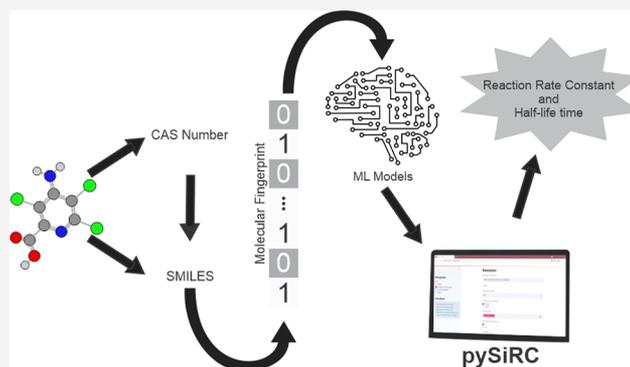
Article Recommendations



Supporting Information

**ABSTRACT:** We developed a web application structured in a machine learning and molecular fingerprint algorithm for the automatic calculation of the reaction rate constant of the oxidative processes of organic pollutants by  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals in the aqueous phase—the pySiRC platform. The model development followed the OECD principles: internal and external validation, applicability domain, and mechanistic interpretation. Three machine learning algorithms combined with molecular fingerprints were evaluated, and all the models resulted in high goodness-of-fit for the training set with  $R^2 > 0.931$  for the  $\bullet\text{OH}$  radical and  $R^2 > 0.916$  for the  $\text{SO}_4^{\bullet-}$  radical and good predictive capacity for the test set with  $R_{\text{ext}}^2 = Q_{\text{ext}}^2$  values in the range of 0.639–0.823 and 0.767–0.824 for the  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals. The model was interpreted using the SHAP (SHapley Additive exPlanations) method: the results showed that the model developed made the prediction based on a reasonable understanding of how electron-withdrawing and -donating groups interfere with the reactivity of the  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals. We hope that our models and web interface can stimulate and expand the application and interpretation of kinetic research on contaminants in water treatment units based on advanced oxidative technologies.

**KEYWORDS:** artificial intelligence, emerging contaminant degradation, kinetic parameters, apps and web applications



## 1. INTRODUCTION

The presence of organic contaminants (OCs) in wastewater has drawn the attention of the scientific community worldwide because most of the water treatment plants were not designed to deal with these emerging pollutants.<sup>1,2</sup> Several studies have shown the negative health impacts caused by the non-removal of these compounds.<sup>3–5</sup> In order to investigate this condition, regulatory agencies such as the U.S. EPA, EU Directive, and Brazilian NR<sup>6–8</sup> have created control standards to oversee public and private companies in water treatment processes. One of the major problems associated with the removal of OCs in water is related to conventional water treatment methods.<sup>9–11</sup> In recent years, advanced oxidation technologies (AOTs) have been employed as a powerful tool to degrade OCs that are not removed by conventional treatments.<sup>11–15</sup> The main AOTs are based on the production of hydroxyl ( $\bullet\text{OH}$ ) and sulfate anion ( $\text{SO}_4^{\bullet-}$ ) radicals, which are strong and highly reactive oxidants capable of reducing or even mineralizing the OCs present in wastewater.<sup>16,17</sup> The main degradation mechanisms associated with the  $\bullet\text{OH}/\text{SO}_4^{\bullet-}$  radicals and OCs are as follows: (i) radical

adduct formation, (ii) hydrogen atom abstraction, and (iii) single electron transfer.<sup>12,18,19</sup>

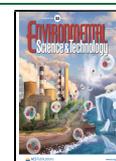
From an experimental point of view, the estimation of the reaction rate constant is considered analytically challenging and requires appreciable costs: experimental measurements of the kinetics of OCs by oxidative processes are considered complex, expensive, and equipment dependent.<sup>19,20</sup> The reaction rate constant reveals important and appropriate criteria to evaluate the efficiency of the degradation of OCs in water through oxidative processes—a high value for this kinetic parameter indicates faster oxidation.<sup>21</sup> Due to the experimental difficulties to describe the reactivity of the OC reaction processes through oxidative processes, the application of theoretical and computational protocols have become imperative for a better under-

Received: June 30, 2021

Revised: August 20, 2021

Accepted: August 23, 2021

Published: September 2, 2021



standing of these phenomena. A series of papers used electronic structure calculations combined with modern reaction rate theories for treating the degradation kinetics of OCs with  $\bullet\text{OH}/\text{SO}_4^{\bullet-}$  radicals, providing measurable parameters, for example, reaction rate constant and half-life time.<sup>14,22–24</sup> However, the theoretical description of the reaction rate constant requires accurate quantum chemical information about the potential energy surface, which makes it a herculean protocol, considering the high computational effort demanded.<sup>25–28</sup>

The practical relevance of the reaction rate constant in evaluating the efficiency of an oxidative process combined with the difficulties associated with experimental measurements and theoretical predictions inspired us to develop a simple but accurate estimation model for a wide range of OCs in the aqueous phase. The main protocol for the prediction of the reaction rate constant is based on the quantitative structure–property/activity relationships (QSPR/QSAR). Several studies have been using QSAR/QSPR models to predict the reaction rate constant of OCs with different types of oxidants such as  $\bullet\text{OH}$ ,<sup>29–31</sup>  $\text{SO}_4^{\bullet-}$ ,<sup>32,33</sup>  $\text{O}_3$ ,<sup>21,34,35</sup> and  $\text{ClO}_2$ ,<sup>36,37</sup> however, the development of QSAR models requires calculations of specific molecular descriptors. Despite the widespread use of QSAR/QSPR models, the selection of suitable molecular descriptors is highly dependent on complex computational quantum chemistry protocols.<sup>38</sup>

Modern methods and the presence of a larger data set have allowed machine learning (ML) algorithms to make increasingly accurate predictions about molecular properties.<sup>38–40</sup> ML has proven to be a powerful tool for competing or even surpassing conventional *ab initio* calculations.<sup>41–43</sup> Although data sets have grown exponentially due to data-driven analysis, ML models require useful and related information from a molecule in a fixed dimension representation.<sup>43</sup> One of the successful protocols to encode molecular structures is the molecular fingerprints (MF) representation<sup>44–47</sup>—which involves transforming a molecular representation into a sequence of binary digits (bits) in order to account for the presence or absence of molecular fragments. Recently, Zhong et al.<sup>38</sup> combined a deep neural network with MF (DNN-MF) to predict the reaction rate constant of OCs with the  $\bullet\text{OH}$  radical in the aqueous phase, obtaining a precision comparable to traditional QSAR models. Given the practical importance of the reaction rate constant, they suggested the following: “To make the established DNN-ML models broadly available, it will be useful to develop APPs or web applications for automatic calculation. Users can simply input the chemical names, CAS numbers, or the chemical structures of new compounds and click the prediction button”. Furthermore, these authors emphasized the importance of developing the model while taking into account the wide availability of computational routines in the python language.

In recent years, the use of web applications has helped in the use of ML models in a practical way.<sup>48–50</sup> Here, our study aimed to develop a free and user-friendly web application based on holistic ML-MF models to calculate the reaction rate constant of the attack of  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals against aqueous OCs, referred to as “pySiRC”, following the OECD principles.<sup>51</sup> To integrate the regulatory process for the development of the QSAR model, this work contains the five principles according to the OECD guidelines:<sup>30,51,52</sup> a defined endpoint; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness-of-fit, robustness and predictive power; and a mechanistic interpretation, if possible.

## 2. MATERIALS AND METHODS

**2.1. Data Sets and MF.** Two databases were built for the reaction rate constants of an oxidative process mediated by  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals: in the former, a group of 1374 OCs ( $k_{\text{OH}}$ ) and in the latter, a group of 400 OCs ( $k_{\text{SO}_4^{\bullet-}}$ ). They were catalogued from the Supporting Information from ref. Zhong et al.,<sup>53</sup> Borhani et al.,<sup>20</sup> Ortiz et al.,<sup>54</sup> Xiao et al.,<sup>32</sup> Wojnárovits and Takács<sup>24</sup> the IscoKin,<sup>55</sup> and NIST<sup>56</sup> database. The kinetic parameters are catalogued under standard conditions, 25 °C and 1 mol·L<sup>-1</sup> in the aqueous phase. The only change made to the catalogued data was to convert the compound names or CAS number to SMILES (simplified molecular-input line-entry system) using CIRpy (<https://github.com/mcs07/CIRpy>). These SMILES were converted into two types of MF—Morgan and MACCS fingerprints—with the RDKit program (<https://www.rdkit.org>). For the Morgan Fingerprint, a scan of the length of the MF was performed using the following values: 512, 1024, 2048, 3072, 4096, and 8192. The average value was used when more than one reaction rate constant was reported for the same OC. The reaction rate constants were scaled into natural logarithm scales and later normalized from 0–1 to reduce the range of values and symmetrize the response.<sup>20,57</sup> The full database is available in the file named “SupInfoDataSet.xlsx” in the Supporting Information file.

**2.2. ML Models and Validation.** Three ML algorithms—Neural Network (NN),<sup>58</sup> Random Forest (RF),<sup>59</sup> and XGBoost<sup>60</sup>—were built to predict the reaction rate constant of the attack of hydroxyl ( $k_{\text{OH}}$ ) and sulfate radicals ( $k_{\text{SO}_4^{\bullet-}}$ ). The models were developed using the scikit-learn packages.<sup>61,62</sup> Internal and external validations were necessary to assess the reliability of ML models and to verify their robustness and predictive capacity.<sup>21</sup>  $k_{\text{OH}}$  and  $k_{\text{SO}_4^{\bullet-}}$  data sets were randomly split into a training set (80%) and a test set (20%). Performance indices—correlation coefficient ( $R^2$ ), Pearson correlation coefficient of prediction ( $r^2$ ), root-mean-square deviation (RMSE), and external validation ( $Q_{\text{ext}}^2$ )—were calculated using the formulas

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{exp}} - y_{\text{pred}})^2}{\sum_{i=1}^n (y_{\text{exp}} - \bar{y}_{\text{exp}})^2} \quad (1)$$

$$r^2 = \frac{\sum_{i=1}^n (y_{\text{exp}} - \bar{y}_{\text{exp}}) \cdot (y_{\text{pred}} - \bar{y}_{\text{pred}})}{\sqrt{\sum_{i=1}^n (y_{\text{exp}} - \bar{y}_{\text{exp}})^2} \sqrt{\sum_{i=1}^n (y_{\text{pred}} - \bar{y}_{\text{pred}})^2}} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_{\text{exp}} - y_{\text{pred}})^2}{n}} \quad (3)$$

$$Q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{exp}} - y_{\text{pred}})^2}{\sum_{i=1}^n (y_{\text{exp}} - \bar{y}_{\text{exp}}^{\text{tr}})^2} \quad (4)$$

where  $y_{\text{exp}}$ ,  $y_{\text{pred}}$ ,  $\bar{y}_{\text{exp}}$ , and  $\bar{y}_{\text{pred}}$  are the experimental, predicted, and average of the experimental and predicted values of the dependent variable (over the validation set), respectively.  $\bar{y}_{\text{exp}}^{\text{tr}}$  is the average value of the dependent variable for the training set—the sums cover all the compounds in the validation set. For the training data set, a 10-fold cross-validation method was used, which randomly divided the data into 10 subgroups. With subsequent optimization cycles, one data set was retained for validation and the others were used for training. External

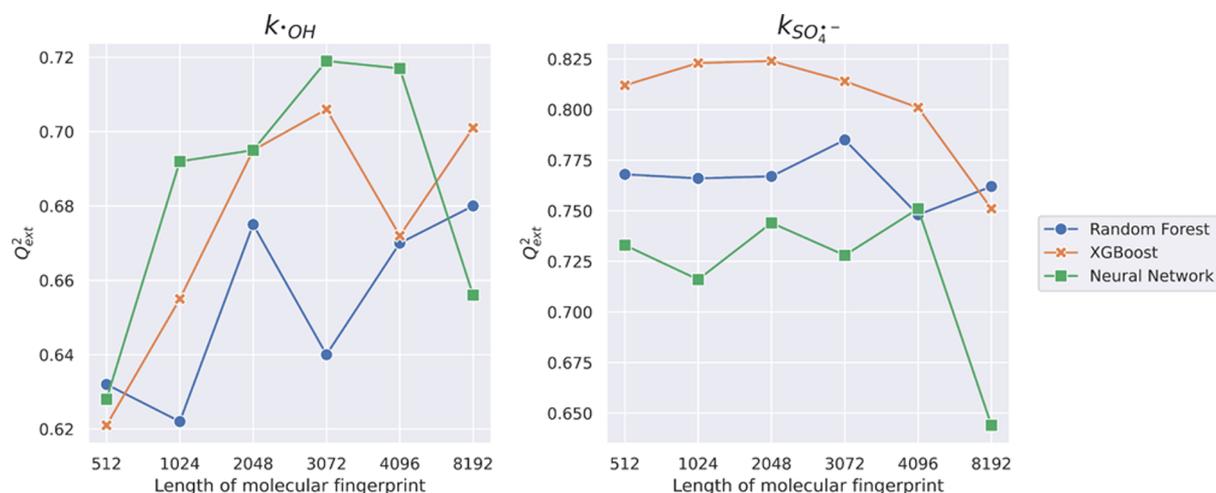


Figure 1. Effects of the length of the Morgan fingerprint on the  $Q_{\text{ext}}^2$  values to RF, XGB, and NN models.

Table 1. Internal and External Validation Parameters of the ML Models Applied to  $k_{\text{OH}}$  and  $k_{\text{SO}_4^-}$  Data Sets with Morgan and MACCS Key Fingerprints

models	Morgan fingerprint									
	training set				test set					
	$\bullet\text{OH}$		$\text{SO}_4^-$		$\bullet\text{OH}$			$\text{SO}_4^-$		
	$R^2$	RMSE	$R^2$	RMSE	$r^2$	$R_{\text{ext}}^2 = Q_{\text{ext}}^2$	RMSE <sub>ext</sub>	$r^2$	$R_{\text{ext}}^2 = Q_{\text{ext}}^2$	RMSE <sub>ext</sub>
NN	0.992	0.011	0.988	0.017	0.851	0.719	0.083	0.871	0.745	0.097
XGB	0.937	0.101	0.992	0.125	0.840	0.707	0.085	0.920	0.824	0.081
RF	0.931	0.099	0.921	0.121	0.801	0.639	0.094	0.897	0.767	0.093
models	MACCS fingerprint									
	training set				test set					
	$\bullet\text{OH}$		$\text{SO}_4^-$		$\bullet\text{OH}$			$\text{SO}_4^-$		
	$R^2$	RMSE	$R^2$	RMSE	$r^2$	$R_{\text{ext}}^2 = Q_{\text{ext}}^2$	RMSE <sub>ext</sub>	$r^2$	$R_{\text{ext}}^2 = Q_{\text{ext}}^2$	RMSE <sub>ext</sub>
NN	0.965	0.008	0.957	0.015	0.897	0.802	0.070	0.886	0.790	0.088
XGB	0.960	0.089	0.978	0.121	0.907	0.823	0.066	0.889	0.799	0.086
RF	0.933	0.090	0.916	0.112	0.905	0.814	0.068	0.899	0.807	0.084

validation (test data set) for ML models was applied, assuming a subgroup excluded from the training set. Furthermore, the  $p$ -value and residual plots were also studied as a validation measure.

**2.3. Applicability Domain.** To assess whether the models developed in this workflow have a reliable prediction, the applicability domain (AD)—crucial for any QSAR/QSPR model<sup>51,63</sup>—was calculated. The AD aims to compare the similarity of the query compound with the compounds in the training data set. If there is a significant similarity, defined by a predefined threshold, the query compound will be reliably predicted by our models. The Tanimoto index<sup>64</sup>  $T_c(A, B)$  was used to evaluate the similarity between two compounds, A and B, for both types of MF and radicals, according to the following equation

$$T_c(A, B) = \frac{c}{a + b - c} \quad (5)$$

where  $a$  and  $b$  are the numbers of structural features, or bits set to 1, in each molecule, and  $c$  is the number in common.

The maximum similarity, which refers to the maximum value of the Tanimoto index between all similarity values obtained, and the mean similarity, which refers to the mean of these similarity values, were used as metrics to assess whether the

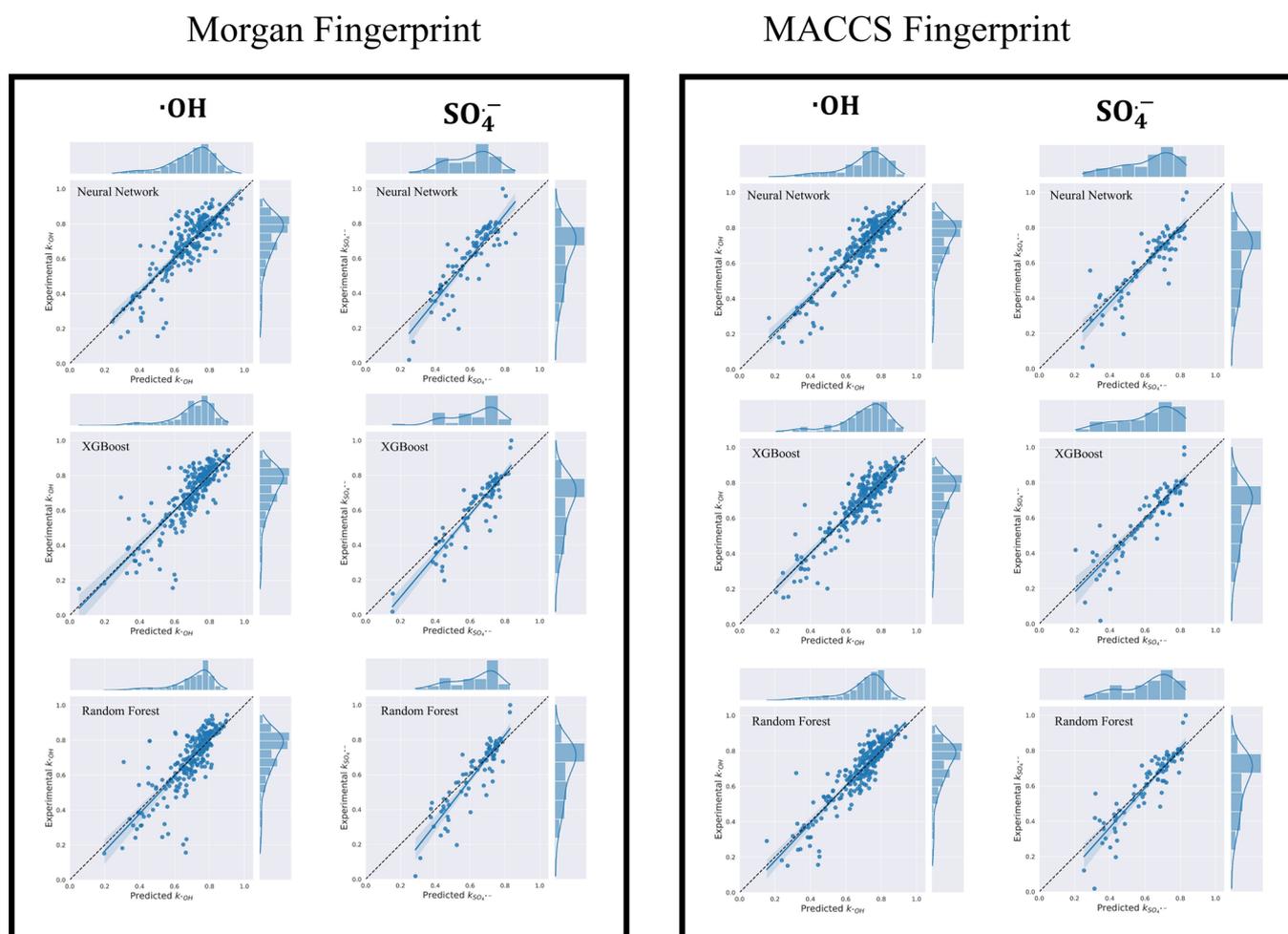
query compound is inside or outside the AD. For each predefined threshold, compounds that were outside the AD were removed from the test data set and RMSE<sub>ext</sub> was recalculated. The optimal limit was the one that obtained the lowest RMSE<sub>ext</sub> value with the lowest possible number of compounds outside the AD.

**2.4. Model Interpretation.** One of the most important validation principles in the development of QSAR/QSPR models is the mechanistic interpretation of the model.<sup>51</sup> It is necessary to perform the prediction based on essential chemical interpretations. In the case of molecular descriptors, there are physico-chemical properties that are chosen to make the best prediction. For the case of MF, the recently developed interpretable explanatory method called SHapley Additive exPlanations (SHAP)<sup>65</sup> was used to rationalize the predictions of the reaction rate constants of the  $\bullet\text{OH}$  and  $\text{SO}_4^-$  radicals.

### 3. RESULTS AND DISCUSSION

**3.1. Effects of the MF Length.** In order to select the best length of the MF for the Morgan type, a preliminary analysis was carried out. In Figure 1, it is possible to observe the effect that the variation in the length of MF causes on the predictive power ( $Q_{\text{ext}}^2$ ) of the different models employed. For the  $\bullet\text{OH}$  radical data set, the length of 3072 bits had the best performance for the

## External Validation



**Figure 2.** Plot of predicted *versus* experimental values for data sets of the oxidative process mediated by  $\bullet\text{OH}$  and  $\text{SO}_4^{\bullet-}$  radicals in aqueous OCs. External validation of the ML models is performed comparing the reference line (dashed) with the linear representation of the data (solid).

XGBoost and NN models, while for the RF model, the length of 8192 bits was the best. In this sense, the length of 3072 bits was selected; lengths greater than 3072 bits led to a negligible increase in  $Q_{\text{ext}}^2$  but made the calculations more demanding. For the radical  $\text{SO}_4^{\bullet-}$ , following the same protocol, the length of 2048 bits was selected.

**3.2. Internal and External Validation.** A generic workflow was fitted with experimental data to train and test three ML models combined with two types of MF of OCs. Two ML ensembles based on decision trees (RF and XGBoost) and Neural Networking were selected.<sup>62</sup> Previous studies showed that these models are appropriate to predict relevant chemical properties.<sup>25,41,43</sup>

For internal validation, statistical criteria— $R^2$  and RMSE—were analyzed to access the performance of the models using the 10-fold cross-validation method (10-fold CV, see Section 2.2). An average of the  $R^2$  and RMSE values of each cross-validation subgroup for the six models developed in this work was used to assess the training performance, and they are shown in Table 1. ML models yield  $R^2$  (RMSE) values in the range of 0.931–0.992 (0.011–0.099) and 0.933–0.965 (0.008–0.090) for  $k_{\bullet\text{OH}}$  data sets with Morgan and MACCS fingerprints, respectively. For  $k_{\text{SO}_4^{\bullet-}}$  data sets, the models yield  $R^2$  (RMSE) values in the range

of 0.921–0.992 (0.017–0.125) and 0.916–0.978 (0.015–0.121) with Morgan and MACCS fingerprints, respectively. In addition, Y-randomization was performed to validate the robustness of the models.<sup>66,67</sup> A new model was developed keeping the original independent variable constant, while the vector of the dependent variable was randomly shuffled—this procedure was repeated three times. The new  $Q_{\text{ext}}^2$  values were lower than in our original model (see Table S2 in Supporting Information file), suggesting that the results from our original model were not accidental.

It is noteworthy that only the values of  $R^2$  and RMSE of the training set did not provide enough statistical criteria to indicate the reliability of the developed models.<sup>63,68,69</sup> Further external validation criteria were applied in this work for both data sets. Before the development of the model, the test set (external validation)—a subgroup representing 20% of the original data set—was randomly selected to perform additional validation criteria of the models employed. Assessment of the performance of external validation was permitted by  $r^2$  and  $Q_{\text{ext}}^2$  statistical parameters (see eqs 2 and 4), which were studied for both data sets.  $r^2$  was calculated to verify the linear correlation between the predicted and experimental values, resulting in  $r^2 > 0.80$  for all models employed (see Table 1). The predictive power of the

**Table 2.** Comparison with Previous Works between Different Models for Internal (Training Set) and External (Test Set) Validation in the Prediction of  $k_{OH}$  and  $k_{SO_4^{\bullet-}}$  Values

model	algorithm	n <sup>a</sup>	•OH data					
			p <sup>b</sup>	training set		test set		
				R <sup>2</sup>	RMSE	R <sup>2</sup> <sub>ext</sub>	RMSE	Q <sup>2</sup> <sub>ext</sub>
Wang et al., 2009	MLR	55	4	0.905	0.139	0.962	0.0079	0.922
Kušić et al., 2009	GA-MLR <sup>c</sup>	78	4	0.735	0.174	0.760	0.200	
Sudhakaran and Amy, 2013	PCA-MLR <sup>d</sup>	83	2	0.918				0.856
Jin et al., 2015	MLR	118	7	0.823	0.204	0.772	0.329	
Borhani et al., 2016	BPSO-MLR <sup>e</sup>	453	8	0.716	0.347	0.724	0.356	0.841
Luo et al., 2017	MLR	526	13	0.805	0.165	0.802	0.232	0.801
Zhong et al., 2020	DNN-MF	593	0	0.972	0.135	0.789	0.329	
Zhong et al., 2021	DNN-MF	1089	0	0.910	0.210	0.600	0.330	
this work	XGB-MACCS	1374	0	0.960	0.089	0.823	0.066	0.823
model	algorithm	n	SO <sub>4</sub> <sup>•-</sup> data					
			p	training set		test set		
				R <sup>2</sup>	RMSE	R <sup>2</sup> <sub>ext</sub>	RMSE	Q <sup>2</sup> <sub>ext</sub>
Xiao et al., 2015	PCA-MLR	85	3	0.87		0.89		0.89
Gupta and Basant, 2016	PCA-DTB <sup>f</sup>	115	3	0.981	0.150	0.975	0.180	
Ye et al., 2017	GCM-MLR <sup>g</sup>	113	32	0.88		0.62		
this work	XGB-Morgan	400	0	0.922	0.125	0.824	0.081	0.824

<sup>a</sup>n = the total number of chemical compounds in the data set. <sup>b</sup>p = the number of molecular descriptors. <sup>c</sup>GA: genetic algorithm; MLR: multiple linear regression. <sup>d</sup>PCA: principal component analysis. <sup>e</sup>BPSO: binary particle swarm optimization. <sup>f</sup>DTB: decision tree boost. <sup>g</sup>group contribution method.

models ( $Q_{ext}^2$ ) was calculated for all the models and resulted in  $Q_{ext}^2$  values in the range of 0.639–0.824, indicating a high goodness-of-fit. In addition, all models showed a  $p$ -value < 0.05, indicating that for the 95% confidence interval,  $r^2$  is statistically significant. Figure S1 (see the Supporting Information file) shows the residual plots of all the models studied and as expected, the residues are randomly distributed around zero. All these results provided a good predictive capacity and clearly indicated that our models excluded overfitted behavior.

The difference between the  $RMSE_{ext}$  and  $Q_{ext}^2$  values of the test set was used to assess the performance of ML models to predict the reaction rate constants. Based on the  $RMSE_{ext}$  and  $Q_{ext}^2$  values, the predictive powers of all the models can be ranked from the best to the worst as follows: (i) NN > XGB > RF for  $k_{OH}$  with Morgan, (ii) XGB > RF > NN for  $k_{OH}$  with MACCS, (iii) XGB > RF > NN for  $k_{SO_4^{\bullet-}}$  with Morgan, and (iv) RF > XGB > NN for  $k_{SO_4^{\bullet-}}$  with MACCS.

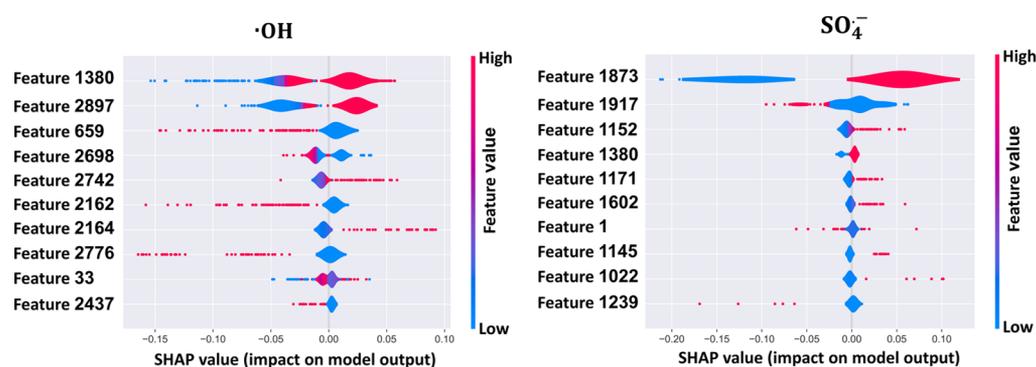
The prediction *versus* experimental plots of the models for both data sets are shown in Figure 2. They described values in a significant range from  $9.51 \times 10^6$ – $3.4 \times 10^{10} M^{-1} s^{-1}$  in the case of  $k_{OH}$  data set and  $1.60 \times 10^4$ – $1.61 \times 10^{10} M^{-1} s^{-1}$  in the case of  $k_{SO_4^{\bullet-}}$  data set. The dashed line in bold is the reference line ( $r^2 = 1$ ), which is parametrized to check the capacity of the prediction observed by ML models: when the data deviate from the reference line, the data are over or under what was predicted; on the other hand, when the data tend to the reference line, they are considered well predicted. All ML models applied to both data sets attained a successful representation (solid line) of the reference line, indicating the robustness of the models used to evaluate external data.

In order to evaluate the prediction of the models in relation to the experimental data (see the histogram in Figure 2), we performed a two-tailed Dunnett's multiple comparison test,<sup>70,71</sup> considering the experimental as a control and the models as different treatments. This test was chosen because it is a

powerful statistic, being able to discover significant differences, even if still relatively small, among groups. At a threshold of  $p < 0.05$ , statistically significant mean differences were not found between the experimental (control) and the models (treatments), highlighting a very similar performance of the optimized models.

**3.3. Comparison with Previous Studies.** To date and to the best of our knowledge, this is the first study that has used ML models combined with MF to predict the rate constant of OCs with a sulfate radical. The sulfate radical has been reported as a modern and efficient oxidizing species regarding the mitigation of OCs. Furthermore,  $SO_4^{\bullet-}$  has showed substantial advantages when compared with the  $OH^{\bullet}$  radical: (i) simplicity in the raw material stock; (ii) weakly affected by organic matter and alkalinity in the presence of water; and (iii) higher quantum yield.<sup>72,73</sup>

In the case of the  $OH^{\bullet}$  radical, we are expanding the previous study performed by Zhong and co-workers.<sup>38,53</sup> Table 2 shows a comparison with previous works which applied ML or QSAR variants. The work of Zhong and co-workers<sup>38,53</sup> is the most recent work and the first to use DNN-MF to predict the reaction rate constant of OC degradation by  $OH^{\bullet}$  radicals. In their first work, they reported  $R^2$  (RMSE) values for the training and test sets close to 0.972 (0.135) and 0.747 (0.329), respectively. On the other hand, in a subsequent work, with the increase of the data set, the  $R_{ext}^2$  value decreased to 0.60. The results of Zhong et al.'s study indicated the good predictive power of the reaction rate constants, which was ratified by the robustness of  $R^2$  and RMSE parameters (see Table 2). Our results showed a similar predictive power, and in some models even better than those obtained by them. To avoid overfitting, the external validation parameter  $Q_{ext}^2 = 0.823$  was estimated, providing a consistency similar to previous works presented in Table 2. However, the works that estimated  $Q_{ext}^2$  parameters in Table 2 are dependent on the molecular descriptor, creating a link between the



**Figure 3.** Importance of the representative MF (top 10) and the SHAP values for the ML models applied to  $k_{\cdot\text{OH}}$  and  $k_{\text{SO}_4^{\cdot-}}$  data sets with Morgan fingerprint based.

**Table 3.** Top-10 Features in the SHAP Values Plot of Figure 3 and Represented Atom Groups and Effects on the Predictions of the Reaction Rate Constant

$\cdot\text{OH}$			$\text{SO}_4^{\cdot-}$		
Feature <sup>a</sup>	Atom group	Effect	Feature	Atom group	Effect
1380		Increase	1873		Increase
2897		Increase	1917		Decrease
659		Decrease	1152		Increase
2698		Decrease	1380		Increase
2742		Increase	1171		Increase
2162		Decrease	1602		Increase
2164		Increase	1		Decrease
2776		Decrease	1145		Increase
33		Increase	1022		Increase
2437		Decrease	1239		Decrease

<sup>a</sup>For the hydroxyl radical, the MF length of 3072 was selected and for the sulfate radical of 2048.

predictive models and the quality of molecular parameters calculated.

**3.4. SHAP.** The prediction of kinetic parameters by artificial intelligence algorithms cannot rely on spurious features of the molecular system. The “understanding process” of ML algorithms must reflect the specificity of the molecular structures involved in the reaction process, providing a molecular comprehension of the major factors that led to the predictions made by the model. Consequently, we apply the SHAP<sup>65</sup> methodology to the XGB model based on the Morgan fingerprint to interpret the relevant structural molecular features

in the process of estimating the reactivity of the hydroxyl and sulfate radicals with organic compounds.

Figure 3 shows the SHAP graph with the 10 main features “learned” by the model as relevant to predict the reactivity of organic pollutants with  $\cdot\text{OH}$  and  $\text{SO}_4^{\cdot-}$  radicals. The blue color represents the absence of a certain feature, that is, a molecular fragment, while the red color indicates the presence of this fragment. The positive or negative SHAP values ( $x$ -axis) of each feature mean that the presence/absence of this fragment can increase or decrease the reaction rate constant.

**Table 4.** Thresholds of Similarity, Number of Compounds Outside the AD for Each Threshold Value, and Corresponding RMSE<sub>ext</sub>

			Morgan fingerprint					
			NN		XGB		RF	
threshold <sup>a</sup>			RMSE <sub>ext</sub>	NCOAD <sup>b</sup>	RMSE <sub>ext</sub>	NCOAD	RMSE <sub>ext</sub>	NCOAD
$k_{OH}^{\bullet}$	maximum	0.100	0.083	0	0.085	1	0.094	0
	mean	0.005	0.080	1	0.081	1	0.090	1
$k_{SO_4^{\bullet-}}$	maximum	0.150	0.097	0	0.081	0	0.093	0
	mean	0.010	0.097	0	0.081	0	0.093	0
			MACCS fingerprint					
			NN		XGB		RF	
threshold			RMSE <sub>ext</sub>	NCOAD	RMSE <sub>ext</sub>	NCOAD	RMSE <sub>ext</sub>	NCOAD
$k_{OH}^{\bullet}$	maximum	0.500	0.068	2	0.065	2	0.065	2
	mean	0.035	0.068	2	0.066	2	0.066	2
$k_{SO_4^{\bullet-}}$	maximum	0.500	0.083	3	0.084	3	0.082	3
	mean	0.080	0.083	4	0.081	4	0.081	4

<sup>a</sup>Threshold  $k_{OH}^{\bullet}$  XGB = 0.150. <sup>b</sup>Number of compounds outside the AD.

Table 3 shows the 10 most important fragments that can affect the reactivity both  $\bullet OH$  and  $SO_4^{\bullet-}$  radicals, the same is shown in Figure 3. For most of the features, when the fragments are absent—represented by the blue dots—the SHAP values are very close to zero, suggesting that these fragments are irrelevant to the predictions of the models.

Our protocol revealed the decisive role of electron-donating groups in increasing the reactivity of  $k_{OH}^{\bullet}$ , while the presence of electron-withdrawing groups decreases their reactivity, which is consistent with the experimental observation that showed a preference for H abstraction in aromatic groups.<sup>74,75</sup> For example, features 1380 and 2897 (aromatic carbon), 2742 (ethylene carbon), and 2164 (sulfur atom) were identified as groups of atoms that increase reactivity and features 659 and 2776 (halogens) and 2668 (carbonyl oxygen) were identified as the group that decreased reactivity.

In the case of the sulfate radical, recent studies have shown that the degradation process can be initiated primarily with an electron transfer of a carbon atom in the aromatic ring and the presence of electron-donating groups increases its reactivity.<sup>32,76</sup> This is consistent with the observation found by our model, in which features 1873 (aromatic carbon), 1152 (secondary amine), 1171 (tertiary amine), and 1602 (hydroxyl) indicated that the presence of these fragments increases the reactivity of the kinetic constant ( $k_{SO_4^{\bullet-}}$ ). Therefore, it is possible to conclude that the ML model made predictions based on a reasonable understanding of how the groups affect the reaction rate constant of the  $\bullet OH$  and  $SO_4^{\bullet-}$  radicals, showing that it is possible to have a chemical interpretation behind an ML protocol.

Furthermore, it is important to emphasize that the AD has been substantially expanded by our model for the study of sulfate radical reactivity. It was added in our database compounds with sulfur (S), phosphorus (P), and iodine atoms (I)—which were not accounted for in previous models<sup>32</sup>—and we were not restricted to predicting aromatic compounds.<sup>76</sup> Another significant advance of the SHAP protocol application is the discrimination of the contribution of the oxygen atom as a function of its chemical bonds, while the previous description based on QSAR only presents the contribution of the ratio of oxygen atoms to carbon atoms (#O:C), our model describes the relevant role of the C=O functional group in decreasing

reactivity and the  $-OH$  group in increasing reactivity. Additionally, the relevance in the degradation process of functional groups with the presence of the nitrogen atom in the molecular structure of OCs is elucidated.

**3.5. Analysis of the AD of the Regression Models.** The analysis of the AD was performed varying different thresholds, from 0 to 0.35 for Morgan and from 0 to 0.5 for MACCS, both with an increase of 0.05 for the maximum threshold. A similar procedure was used with the mean threshold, ranging from 0 to 0.095 with an increase of 0.005. Table 4 shows the optimal values found for the AD for the two types of MF and for the  $\bullet OH$  and  $SO_4^{\bullet-}$  radicals. A threshold of 0.100 led to the lowest RMSE<sub>ext</sub> value for the NN and RF models and a limit of 0.150 for the XGB model using Morgan. For the MACCS, a limit of 0.500 yield the lowest RMSE<sub>ext</sub> values for the three models developed. The maximum similarity metric (see Section 2.3) was chosen to indicate whether the compound is within or outside the AD, that is, if the query compound has a value less than the maximum similarity threshold value, the compound will be outside the AD, otherwise, the compound will be within the AD of the model developed.

**3.6. “pySiRC” Platform.** In recent years, the development of ML models applied in different fields of science has grown exponentially. However, the vast majority of these protocols can only be used by specialists—data scientists. The need to be an expert in programming and ML techniques limits the access of many researchers: this was evidenced by Zhong<sup>38</sup> in his article when he said that “This is probably part of the reason that although numerous QSARs have been developed and available for a few decades, they are mostly used by a small group of researchers”. Therefore, we believe that the development of a web application goes beyond an engineering application because it connects the user without prior knowledge in such techniques with the required information in a few clicks. Accessing information in a friendly way for a non-specialist can broaden the interpretation of fundamental processes in several areas, reinforcing the evolution of knowledge scientifically. Below, it is presented in details the protocols to manipulate the pySiRC platform.

The reaction rate constant reveals the efficiency of oxidative processes of contaminants in the aqueous environment, and its measurement or prevision are considered challenging. There-

**Navigation:**

Go to:

- HOME
- Simulator rate constants
- Simulator half-life
- About

**Contribute**

This web platform is free and open-source and you are very welcome to contribute. This Application has GNU GPL license. All source code can be accessed [here](#)

**Simulator**

Type SMILES or CAS Number

Show

Choose a radical reaction

Choose type molecular fingerprint

Morgan

MACCS

Choose ML Models

Return Reaction Rate Constant as

k

ln k

Generate

**Simulator of half-life (OH)**

Type SMILES or CAS Number

Show

Choose

Automatic

Manual

Simulate

**About**

pySiRC was designed to facilitate the automatic researchers to prediction and analysis of reaction rate constants using Machine Learning models, since the theoretical calculation and experimental measures of these kinetic parameters can be extremely challenging. The tools are designed to be quick and simple users with few clicks can make the prediction. The tools are rigorous enough to have applications in academic research. All the source code for this application and the steps for local installation are at <https://github.com/jeffrichardchemistry/pySiRC>. You are free to use, change and distribute this application under the GNU GPL license. The maintainers of this package so far are: Jefferson Richard Dias ([richardquimica@gmail.com](mailto:richardquimica@gmail.com)) and Flavio Olimpio ([flavio\\_olimpio@outlook.com](mailto:flavio_olimpio@outlook.com)).

**Supporters**

FAPEG, CAPES, UnB, UFG, Universidade Estadual de Goiás, LaCiQ, LQTEA

Figure 4. Web main screen of the pySiRC user view.

fore, it is desirable that simple predictive models can be developed for a wider audience, without the need for specific knowledge of chemistry or computational methods of the electronic structure. In this context, to share our model with chemists, pharmacists, and environmental engineers working in water treatment plants, we developed pySiRC (python Simulator of Rate Constant). Additionally, pySiRC was developed to support professional non-experts in chemical kinetics and theoretical and computational chemistry, providing the reactivity profile of OCs. The hosted web graphical platform allows the calculation of the reaction rate constant of the oxidative process of OCs mediated by the  $\cdot\text{OH}$  and  $\text{SO}_4^{\cdot-}$  radicals in the aqueous phase using only a SMILES or CAS Number. pySiRC provides a free and easy user-friendly graphical interface allowing quick analysis of  $k_{\text{OH}}$  and  $k_{\text{SO}_4^{\cdot-}}$ ; the data are in

$\text{M}^{-1} \text{s}^{-1}$ . It can be accessed from [www.pysirc.com.br](http://www.pysirc.com.br). To the best of our knowledge, this is the first free-web application that estimates the reaction rate constant using an automated ML protocol. Figure 4 shows a screenshot of the pySiRC user view.

On the main page of the web application, the user can choose between four options: (i) home; (ii) simulator rate constants; (iii) simulator half-life; and (iv) about. If the “HOME” option is chosen, a description of the application of pySiRC will be displayed. If the “simulator rate constant” option is chosen, a page will be displayed for the user to estimate the reaction rate constant of the desired compound. Here, it is necessary to provide the SMILES or CAS number of the compound, the radical species ( $\cdot\text{OH}$  or  $\text{SO}_4^{\cdot-}$ ), and to select the ML model to perform the prediction. With the SMILES or CAS number, it is

possible to show the 2D molecular structure to check the molecule selected.

From the data of the total reaction rate constant of OCs with the  $\bullet\text{OH}$  radical obtained by ML models, it is possible to calculate the half-life time using  $t_{1/2} = \ln 2 / (k_{\bullet\text{OH}} \times [\bullet\text{OH}])$ , where  $\bullet\text{OH}$  is the concentration of  $\bullet\text{OH}$  radicals in the aqueous media. Therefore, if the “simulator half-life” option is chosen, it is possible to make the prediction of the half-life time automatically, it is necessary to provide the SMILES or CAS number of the compound or manually providing the value of the reaction rate constant—NN model will be used to calculate the reaction rate constant. The half-life time of the reaction is reported at a temperature of 298.15 K, and  $[\bullet\text{OH}] 10^{-15}$ – $10^{-18}$  mol L $^{-1}$ , which usually represents the values found in surface waters.<sup>22,77–79</sup> A plot and a table of the half-life with OH radical concentrations—usually found in the aqueous environment ( $\sim 10^{-15}$ – $10^{-18}$  mol L $^{-1}$ )—are also provided. Finally, if the “about” option is chosen, a brief description of the tools used in the logos of the institutions and funding agencies will be displayed. To permit the use of the web platform and make it widely available, we additionally provide a video simulating the use of the applications that it is in the [Supporting Information](#).

**3.7. Kinetic Perspective.** Water pollution by OCs has become a topic that has demanded great research efforts. Consequently, the evolution of computational protocols assumed an indispensable tool for evaluating the degradation processes involved in the reactional dynamics of these contaminants and consequently reducing the costs and hazards involved in these analyses. However, providing dynamic and kinetic information on contaminant degradation processes is limited by one of the major theoretical bottlenecks in the physicochemical field: application of rate theories from high-level electronic structure methods as anticipated in the seminal article by Professor Henry F. Schaefer III.<sup>80</sup>

Recently, the advance of rate theories in chemical kinetics to predict accurate kinetic parameters from first-principles models gain a strong ally, artificial intelligence protocols. The former created the fundamental bases for the compression of reactive dynamics at the molecular level and the latter leverage technological development. The application of first-principles methods in chemical and physical processes has acquired enormous maturity, enabling the description of intricate problems, such as quantum effects (e.g., tunneling and resonance) in atomic and molecular systems, stochastic motion of particles in a condensed environment, and non-equilibrium effects in classical and quantum formulations.<sup>81,82</sup> However, the advent of modern experimental and computational techniques provides access to kinetic parameters of systems with huge molecular complexity which would demand long periods of time and high-cost resources. Accordingly, the ML predictions paved the way for modeling physical and chemical processes, enriching again the fruitful and long-lasting partnership between artificial intelligence and first-principles models.<sup>27,83</sup>

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.1c04326>.

SMILES data (XLSX)

pySiRC presentation (TXT)

Detailed information on the data set, three ML algorithms, model evaluation process, and model prediction process (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Flávio Olímpio Sanches-Neto – Instituto de Química, Universidade de Brasília, Brasília 70904-970, Brazil; [orcid.org/0000-0002-0664-171X](https://orcid.org/0000-0002-0664-171X); Email: [flavio\\_olimpio@outlook.com](mailto:flavio_olimpio@outlook.com)

Jefferson Richard Dias-Silva – Instituto de Química, Universidade Federal de Goiás, Goiânia 74001-970 Goiás, Brazil; Email: [jrichardquimica@gmail.com](mailto:jrichardquimica@gmail.com)

Luiz Henrique Keng Queiroz Junior – Instituto de Química, Universidade Federal de Goiás, Goiânia 74001-970 Goiás, Brazil; Email: [professorkeng@gmail.com](mailto:professorkeng@gmail.com)

Valter Henrique Carvalho-Silva – Instituto de Química, Universidade de Brasília, Brasília 70904-970, Brazil; Modeling of Physical and Chemical Transformations Division, Theoretical and Structural Chemistry Group, Goiás State University, Anápolis 75132-903, Brazil; [orcid.org/0000-0002-7411-0099](https://orcid.org/0000-0002-7411-0099); Email: [fatioleg@gmail.com](mailto:fatioleg@gmail.com)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.est.1c04326>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge the Brazilian funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support. Valter H. Carvalho-Silva thanks Goiana agency FAPEG for the research funding programs (AUXÍLIO À PESQUISA COLABORATIVA FAPEG-FAPESP/GSP2019011000037).

## ■ REFERENCES

- (1) Cardoso, L. P.; Valim, J. B. Study of Acids Herbicides Removal by Calcined Mg-Al-CO<sub>3</sub>-LDH. *J. Phys. Chem. Solids* **2006**, *67*, 987–993.
- (2) Bao, Y.; Zhang, C.; Yang, W.; Hu, J.; Sun, X. Mechanism and Kinetics Study on the OH-Initiated Oxidation of Organophosphorus Pesticide Trichlorfon in Atmosphere. *Sci. Total Environ.* **2012**, *419*, 144–150.
- (3) Lin, C.; Zhang, L.; Zhang, H.; Wang, Q.; Zhu, J.; Wang, J.; Qian, M. Enantioselective Degradation of Myclobutanil and Famoxadone in Grape. *Environ. Sci. Pollut. Res.* **2018**, *25*, 2718–2725.
- (4) Organization, W. H.; et al. *Public Health Impact of Pesticides Used in Agriculture*; World Health Organization, 1990.
- (5) Lofrano, G.; Pedrazzani, R.; Libralato, G.; Carotenuto, M. Advanced Oxidation Processes for Antibiotics Removal: A Review. *Curr. Org. Chem.* **2017**, *21*, 1054–1067.
- (6) Bartolomeu, M.; Neves, M. G. P. M. S.; Faustino, M. A. F.; Almeida, A. Wastewater Chemical Contaminants: Remediation by Advanced Oxidation Processes. *Photochem. Photobiol. Sci.* **2018**, *17*, 1573–1598.
- (7) Sanches, S.; Barreto Crespo, M. T.; Pereira, V. J. Drinking Water Treatment of Priority Pesticides Using Low Pressure UV Photolysis and Advanced Oxidation Processes. *Water Res.* **2010**, *44*, 1809–1818.
- (8) Marinho, B. A.; Cristóvão, R. O.; Boaventura, R. A. R.; Vilar, V. J. P. As (III) and Cr (VI) Oxyanion Removal from Water by Advanced Oxidation/Reduction Processes—a Review. *Environ. Sci. Pollut. Res.* **2019**, *26*, 2203–2227.

- (9) Almasi, A.; Dargahi, A.; Amrane, A.; Fazlzadeh, M.; Mahmoudi, M.; Hashemian, A. H.; et al. Effect of the Retention Time and the Phenol Concentration on the Stabilization Pond Efficiency in the Treatment of Oil Refinery Wastewater. *Fresenius Environ. Bull.* **2014**, *23*, 2541–2548.
- (10) Ding, H.; Wu, Y.; Zou, B.; Lou, Q.; Zhang, W.; Zhong, J.; Lu, L.; Dai, G. Simultaneous Removal and Degradation Characteristics of Sulfonamide, Tetracycline, and Quinolone Antibiotics by Laccase-Mediated Oxidation Coupled with Soil Adsorption. *J. Hazard. Mater.* **2016**, *307*, 350–358.
- (11) Kurt, A.; Mert, B. K.; Özençin, N.; Sivrioğlu, Ö.; Yonar, T. Treatment of Antibiotics in Wastewater Using Advanced Oxidation Processes (AOPs). *Physico-Chemical Wastewater Treatment and Resource Recovery*; IntechOpen, 2017; p 175.
- (12) Mei, Q.; Sun, J.; Han, D.; Wei, B.; An, Z.; Wang, X.; Xie, J.; Zhan, J.; He, M. Sulfate and Hydroxyl Radicals-Initiated Degradation Reaction on Phenolic Contaminants in the Aqueous Phase: Mechanisms, Kinetics and Toxicity Assessment. *Chem. Eng. J.* **2019**, *373*, 668–676.
- (13) Graça, C. A. L.; Maniero, M. G.; De Andrade, L. M.; Roberto Guimarães, J.; Teixeira, A. C. S. C. Evaluation of Amicarbazone Toxicity Removal through Degradation Processes Based on Hydroxyl and Sulfate Radicals. *J. Environ. Sci. Health, Part A: Toxic/Hazard. Subst. Environ. Eng.* **2019**, *54*, 1126–1143.
- (14) Yang, Z.; Su, R.; Luo, S.; Spinney, R.; Cai, M.; Xiao, R.; Wei, Z. Comparison of the Reactivity of Ibuprofen with Sulfate and Hydroxyl Radicals: An Experimental and Theoretical Study. *Sci. Total Environ.* **2017**, *590–591*, 751–760.
- (15) Luo, S.; Gao, L.; Wei, Z.; Spinney, R.; Dionysiou, D. D.; Hu, W.-P.; Chai, L.; Xiao, R. Kinetic and Mechanistic Aspects of Hydroxyl Radical-mediated Degradation of Naproxen and Reaction Intermediates. *Water Res.* **2018**, *137*, 233–241.
- (16) Manonmani, G.; Sandhiya, L.; Senthilkumar, K. Mechanism and Kinetics of Diuron Oxidation Initiated by Hydroxyl Radical: Hydrogen and Chlorine Atom Abstraction Reactions. *J. Phys. Chem. A* **2019**, *123*, 8954–8967.
- (17) Manonmani, G.; Sandhiya, L.; Senthilkumar, K. Mechanism and Kinetics of Diuron Oxidation by Hydroxyl Radical Addition Reaction. *Environ. Sci. Pollut. Res.* **2020**, *27*, 12080–12095.
- (18) Milenković, D. A.; Dimić, D. S.; Avdović, E. H.; Amić, A. D.; Dimitrić Marković, J. M.; Marković, Z. S. Advanced Oxidation Process of Coumarins by Hydroxyl Radical: Towards the New Mechanism Leading to Less Toxic Products. *Chem. Eng. J.* **2020**, *395*, 124971.
- (19) An, T.; Gao, Y.; Li, G.; Kamat, P. V.; Peller, J.; Joyce, M. V. Kinetics and Mechanism of  $\bullet\text{OH}$  Mediated Degradation of Dimethyl Phthalate in Aqueous Solution: Experimental and Theoretical Studies. *Environ. Sci. Technol.* **2014**, *48*, 641–648.
- (20) Borhani, T. N. G.; Saniedanesh, M.; Bagheri, M.; Lim, J. S. QSPR Prediction of the Hydroxyl Radical Rate Constant of Water Contaminants. *Water Res.* **2016**, *98*, 344–353.
- (21) Sudhakaran, S.; Amy, G. L. QSAR Models for Oxidation of Organic Micropollutants in Water Based on Ozone and Hydroxyl Radical Rate Constants and Their Chemical Classification. *Water Res.* **2013**, *47*, 1111–1122.
- (22) Yang, J.; Wang, Z.; Lv, G.; Liu, W.; Wang, Y.; Sun, X.; Gao, J. Indirect Photodegradation of Fludioxonil by Hydroxyl Radical and Singlet Oxygen in Aquatic Environment: Mechanism, Photoproducts Formation and Eco-Toxicity Assessment. *Ecotoxicol. Environ. Saf.* **2020**, *197*, 110644.
- (23) Gao, Y.; Ji, Y.; Li, G.; An, T. Mechanism, Kinetics and Toxicity Assessment of OH-Initiated Transformation of Triclosan in Aquatic Environments. *Water Res.* **2014**, *49*, 360–370.
- (24) Wojnárovits, L.; Takács, E. Rate Constants of Sulfate Radical Anion Reactions with Organic Molecules: A Review. *Chemosphere* **2019**, *220*, 1014–1032.
- (25) Schmitz, G.; Godtlielsen, I. H.; Christiansen, O. Machine Learning for Potential Energy Surfaces: An Extensive Database and Assessment of Methods. *J. Chem. Phys.* **2019**, *150*, 244113.
- (26) Sanches-Neto, F. O.; Coutinho, N. D.; Carvalho-Silva, V. H. A Novel Assessment of the Role of the Methyl Radical and Water Formation Channel in the  $\text{CH}_3\text{OH} + \text{H}$  Reaction. *Phys. Chem. Chem. Phys.* **2017**, *19*, 24467–24477.
- (27) Machado, H. G.; Sanches-Neto, F. O.; Coutinho, N. D.; Mundim, K. C.; Palazzetti, F.; Carvalho-Silva, V. H. “Transitivity”: A Code for Computing Kinetic and Related Parameters in Chemical Transformations and Transport Phenomena. *Molecules* **2019**, *24*, 3478.
- (28) Sanches-Neto, F. O.; Coutinho, N. D.; Palazzetti, F.; Carvalho-Silva, V. H. Temperature Dependence of Rate Constants for the  $\text{H(D)} + \text{CH}_4$  Reaction in Gas and Aqueous Phase: Deformed Transition-State Theory Study Including Quantum Tunneling and Diffusion Effects. *Struct. Chem.* **2020**, *31*, 609–617.
- (29) Yang, Z.; Luo, S.; Wei, Z.; Ye, T.; Spinney, R.; Chen, D.; Xiao, R. Rate Constants of Hydroxyl Radical Oxidation of Polychlorinated Biphenyls in the Gas Phase: A Single-Descriptor Based QSAR and DFT Study. *Environ. Pollut.* **2016**, *211*, 157–164.
- (30) Wang, Y.-n.; Chen, J.; Li, X.; Wang, B.; Cai, X.; Huang, L. Predicting Rate Constants of Hydroxyl Radical Reactions with Organic Pollutants: Algorithm, Validation, Applicability Domain, and Mechanistic Interpretation. *Atmos. Environ.* **2009**, *43*, 1131–1135.
- (31) Kušić, H.; Rasulev, B.; Leszczynska, D.; Leszczynski, J.; Koprivanac, N. Prediction of Rate Constants for Radical Degradation of Aromatic Pollutants in Water Matrix: A QSAR Study. *Chemosphere* **2009**, *75*, 1128–1134.
- (32) Xiao, R.; Ye, T.; Wei, Z.; Luo, S.; Yang, Z.; Spinney, R. Quantitative Structure–Activity Relationship (QSAR) for the Oxidation of Trace Organic Contaminants by Sulfate Radical. *Environ. Sci. Technol.* **2015**, *49*, 13394–13402.
- (33) Gupta, S.; Basant, N. Modeling the Reactivity of Ozone and Sulphate Radicals towards Organic Chemicals in Water Using Machine Learning Approaches. *RSC Adv.* **2016**, *6*, 108448–108457.
- (34) Jiang, J. L.; Yue, X. A.; Chen, Q. F.; Gao, Z. Determination of Ozonation Reaction Rate Constants of Aromatic Pollutants and QSAR Study. *Bull. Environ. Contam. Toxicol.* **2010**, *85*, 568–572.
- (35) Liu, H.; Tan, J.; Yu, H. X.; Liu, H. X.; Wang, L. S.; Wang, Z. Y. Determination of the Apparent Reaction Rate Constants for Ozone Degradation of Substituted Phenols and QSPR/QSAR Analysis. *Int. J. Environ. Res.* **2010**, *4*, 507–512.
- (36) Tratnyek, P. G.; Hoigné, J. Kinetics of Reactions of Chlorine Dioxide (OCIO) in Water—II. Quantitative Structure–Activity Relationships for Phenolic Compounds. *Water Res.* **1994**, *28*, 57–66.
- (37) Lee, Y.; Von Gunten, U. Quantitative Structure–Activity Relationships (QSARs) for the Transformation of Organic Micropollutants during Oxidative Water Treatment. *Water Res.* **2012**, *46*, 6177–6195.
- (38) Zhong, S.; Hu, J.; Fan, X.; Yu, X.; Zhang, H. A Deep Neural Network Combined with Molecular Fingerprints (DNN-MF) to Develop Predictive Models for Hydroxyl Radical Rate Constants of Water Contaminants. *J. Hazard. Mater.* **2020**, *383*, 121141.
- (39) Jha, A.; Chandrasekaran, A.; Kim, C.; Ramprasad, R. Impact of Dataset Uncertainties on Machine Learning Model Predictions: The Example of Polymer Glass Transition Temperatures. *Modell. Simul. Mater. Sci. Eng.* **2019**, *27*, 24002.
- (40) Bapst, V.; Keck, T.; Grabska-Barwińska, A.; Donner, C.; Cubuk, E. D.; Schoenholz, S. S.; Obika, A.; Nelson, A. W. R.; Back, T.; Hassabis, D.; Kohli, P. Unveiling the Predictive Power of Static Structure in Glassy Systems. *Nat. Phys.* **2020**, *16*, 448–454.
- (41) Houston, P. L.; Nandi, A.; Bowman, J. M. A Machine Learning Approach for Prediction of Rate Constants. *J. Phys. Chem. Lett.* **2019**, *10*, S250–S258.
- (42) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Machine Learning of Molecular Electronic Properties in Chemical Compound Space. *New J. Phys.* **2013**, *15*, 95003.
- (43) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9*, 513–530.

- (44) Pilania, G.; Iverson, C. N.; Lookman, T.; Marrone, B. L. Machine-Learning-Based Predictive Modeling of Glass Transition Temperatures: A Case of Polyhydroxyalkanoate Homopolymers and Copolymers. *J. Chem. Inf. Model.* **2019**, *59*, 5013–5025.
- (45) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular Fingerprint Similarity Search in Virtual Screening. *Methods* **2015**, *71*, 58–63.
- (46) Liu, R.; Zhou, D. Using Molecular Fingerprint as Descriptors in the QSPR Study of Lipophilicity. *J. Chem. Inf. Model.* **2008**, *48*, 542–549.
- (47) Steffen, A.; Kogej, T.; Tyrchan, C.; Engkvist, O. Comparison of Molecular Fingerprint Methods on the Basis of Biological Profile Data. *J. Chem. Inf. Model.* **2009**, *49*, 338–347.
- (48) Pogodin, P. V.; Lagunin, A. A.; Rudik, A. V.; Druzhilovskiy, D. S.; Filimonov, D. A.; Poroikov, V. V. AntiBac-Pred: A Web Application for Predicting Antibacterial Activity of Chemical Compounds. *J. Chem. Inf. Model.* **2019**, *59*, 4513–4518.
- (49) Rudik, A. V.; Bezhentsev, V. M.; Dmitriev, A. V.; Druzhilovskiy, D. S.; Lagunin, A. A.; Filimonov, D. A.; Poroikov, V. V. MetaTox: Web Application for Predicting Structure and Toxicity of Xenobiotics' Metabolites. *J. Chem. Inf. Model.* **2017**, *57*, 638–642.
- (50) Braga, R. C.; Alves, V. M.; Muratov, E. N.; Strickland, J.; Kleinstreuer, N.; Tropscha, A.; Andrade, C. H. Pred-Skin: A Fast and Reliable Web Application to Assess Skin Sensitization Effect of Chemicals. *J. Chem. Inf. Model.* **2017**, *57*, 1013–1017.
- (51) OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; Organisation for Economic Co-operation and Development: Paris, France, 2007.
- (52) Tunkel, J.; Mayo, K.; Austin, C.; Hickerson, A.; Howard, P. Practical Considerations on the Use of Predictive Models for Regulatory Purposes. *Environ. Sci. Technol.* **2005**, *39*, 2188–2199.
- (53) Zhong, S.; Zhang, K.; Wang, D.; Zhang, H. Shedding Light on “Black Box” Machine Learning Models for Predicting the Reactivity of HO[Rad] Radicals toward Organic Compounds. *Chem. Eng. J.* **2021**, *405*, 126627.
- (54) Ortiz, E. V.; Bennardi, D. O.; Bacelo, D. E.; Fiorelli, S. E.; Duchowicz, P. R. The Conformation-Independent QSPR Approach for Predicting the Oxidation Rate Constant of Water Micropollutants. *Environ. Sci. Pollut. Res.* **2017**, *24*, 27366–27375.
- (55) Tratnyek, P. G.; Waldemer, R. H.; Powell, J. S. *IscoKin Database of Rate Constants for Reaction of Organic Contaminants with the Major Oxidants Relevant to In Situ Chemical Oxidation*, 2019.
- (56) Huie, R. E. *NDRL/NIST Solution Kinetics Database on the Web*, 2003.
- (57) Sudhakaran, S.; Amy, G. L. QSAR Models for Oxidation of Organic Micropollutants in Water Based on Ozone and Hydroxyl Radical Rate Constants and Their Chemical Classification. *Water Res.* **2013**, *47*, 1111–1122.
- (58) Hansen, L. K.; Salamon, P. Neural Network Ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 993–1001.
- (59) Ho, T. K. Random Decision Forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995; Vol. 1, pp 278–282.
- (60) Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016; pp 785–794.
- (61) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (62) Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, 2019.
- (63) Gramatica, P. External Evaluation of QSAR Models, in Addition to Cross-Validation: Verification of Predictive Capability on Totally New Chemicals. *Mol. Inf.* **2014**, *33*, 311–314.
- (64) Tanimoto, T. T. *Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corp., 1958.
- (65) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67.
- (66) Fan, D.; Yang, H.; Li, F.; Sun, L.; Di, P.; Li, W.; Tang, Y.; Liu, G. In Silico Prediction of Chemical Genotoxicity Using Machine Learning Methods and Structural Alerts. *Toxicol. Res.* **2018**, *7*, 211–220.
- (67) Asadollahi, T.; Dadfarnia, S.; Shabani, A. M. H.; Ghasemi, J. B.; Sarkhosh, M. QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using in Silico Virtual Screening. *Molecules* **2011**, *16*, 1928–1955.
- (68) Kiralj, R.; Ferreira, M. M. C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.* **2009**, *20*, 770–787.
- (69) Roy, K.; Ambure, P.; Aher, R. B. How Important Is to Detect Systematic Error in Predictions and Understand Statistical Applicability Domain of QSAR Models? *Chemom. Intell. Lab. Syst.* **2017**, *162*, 44–54.
- (70) Dunnett, C. W. Multiple Comparison Procedure for Comparing Several Treatments with a Control. *J. Am. Stat. Assoc.* **1955**, *50*, 1096–1121.
- (71) Dunnett, C. W. New Tables for Multiple Comparisons with a Control. *Biometrics* **1964**, *20*, 482–491.
- (72) Ye, T.; Wei, Z.; Spinney, R.; Tang, C.-J.; Luo, S.; Xiao, R.; Dionysiou, D. D. Chemical Structure-Based Predictive Model for the Oxidation of Trace Organic Contaminants by Sulfate Radical. *Water Res.* **2017**, *116*, 106–115.
- (73) He, X.; de la Cruz, A. A.; Dionysiou, D. D. Destruction of Cyanobacterial Toxin Cylindrospermopsin by Hydroxyl Radicals and Sulfate Radicals Using UV-254 Nm Activation of Hydrogen Peroxide, Persulfate and Peroxymonosulfate. *J. Photochem. Photobiol., A* **2013**, *251*, 160–166.
- (74) Minakata, D.; Li, K.; Westerhoff, P.; Crittenden, J. Development of a Group Contribution Method to Predict Aqueous Phase Hydroxyl Radical (HO•) Reaction Rate Constants. *Environ. Sci. Technol.* **2009**, *43*, 6220–6227.
- (75) Minakata, D.; Song, W.; Mezyk, S. P.; Cooper, W. J. Experimental and Theoretical Studies on Aqueous-Phase Reactivity of Hydroxyl Radicals with Multiple Carboxylated and Hydroxylated Benzene Compounds. *Phys. Chem. Chem. Phys.* **2015**, *17*, 11796–11812.
- (76) Luo, S.; Wei, Z.; Dionysiou, D. D.; Spinney, R.; Hu, W.-P.; Chai, L.; Yang, Z.; Ye, T.; Xiao, R. Mechanistic Insight into Reactivity of Sulfate Radical with Aromatic Contaminants through Single-Electron Transfer Pathway. *Chem. Eng. J.* **2017**, *327*, 1056–1065.
- (77) Brezonik, P. L.; Fulkerson-Brekken, J. Nitrate-Induced Photolysis in Natural Waters: Controls on Concentrations of Hydroxyl Radical Photo-Intermediates by Natural Scavenging Agents. *Environ. Sci. Technol.* **1998**, *32*, 3004–3010.
- (78) Burns, J. M.; Cooper, W. J.; Ferry, J. L.; King, D. W.; DiMento, B. P.; McNeill, K.; Miller, C. J.; Miller, W. L.; Peake, B. M.; Rusak, S. A.; Rose, A. L.; Waite, T. D. Methods for Reactive Oxygen Species (ROS) Detection in Aqueous Environments. *Aquat. Sci.* **2012**, *74*, 683–734.
- (79) Sanches-Neto, F. O.; Ramos, B.; Lastre-Acosta, A. M.; Teixeira, A. C. S. C.; Carvalho-Silva, V. H. Aqueous Picloram Degradation by Hydroxyl Radicals: Unveiling Mechanism, Kinetics, and Ecotoxicity through Experimental and Theoretical Approaches. *Chemosphere* **2021**, *278*, 130401.
- (80) Schaefer, H. F., III The Atomic Fluorine + Molecular Hydrogen Potential Energy Surface: The Ecstasy and the Agony. *J. Phys. Chem.* **1985**, *89*, 5336–5343.
- (81) Carvalho-Silva, V. H.; Coutinho, N. D.; Aquilanti, V. Temperature Dependence of Rate Processes beyond Arrhenius and Eyring: Activation and Transitivity. *Front. Chem.* **2019**, *7*, 380.
- (82) Carvalho-Silva, V. H.; Coutinho, N. D.; Aquilanti, V. From the Kinetic Theory of Gases to the Kinetics of Rate Processes: On the Verge of the Thermodynamic and Kinetic Limits. *Molecules* **2020**, *25*, 2098.
- (83) Aquilanti, V.; Coutinho, N. D.; Carvalho-Silva, V. H. Kinetics of Low-Temperature Transitions and Reaction Rate Theory from Non-

Equilibrium Distributions. *Philos. Trans. R. Soc., A* **2017**, *375*, 20160204.

### **PAPER 3:**

SANCHES-NETO, F. O., DIAS-SILVA, J. R., DE OLIVEIRA, V. M., AQUILANTI, V., & CARVALHO-SILVA, V. H. (2022). **Evaluating and elucidating the reactivity of OH radicals with atmospheric organic pollutants: Reaction kinetics and mechanisms by machine learning.** *Atmospheric Environment*, 275, 119019.

In a similar way to the aqueous organic pollutants presented in the previous work, the presence of atmospheric organic pollutants has also gained attention on the world stage, since the presence of these pollutants in the atmosphere has been related to global climate change.

In this sense, we expand the methodology of the previous study, that is, we use a machine learning approach combined with molecular fingerprints to predict the rate constant of atmospheric organic pollutants through hydroxyl radical attack. Additionally, let us stress the mechanistic details revealed by the black box of the developed machine learning models.

My contribution in this work was the development of machine learning models and the writing of the article. The authors Jefferson Richard and Vitor Mendes contributed richly in the creation of the models and in the discussions. Professors Vincenzo Aquilanti and Valter Carvaho helped to correct and supervise the work carried out.



RightsLink



Home



Help ▾



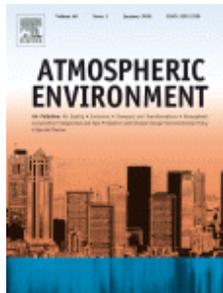
Live Chat



Sign in



Create Account



## Evaluating and elucidating the reactivity of OH radicals with atmospheric organic pollutants: Reaction kinetics and mechanisms by machine learning

**Author:**

Flávio O. Sanches-Neto, Jefferson R. Dias-Silva, Vitor M. de Oliveira, Vincenzo Aquilanti, Valter H. Carvalho-Silva

**Publication:** Atmospheric Environment

**Publisher:** Elsevier

**Date:** 15 April 2022

© 2022 Elsevier Ltd. All rights reserved.

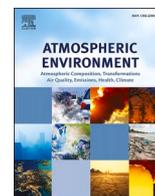
### Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)  
| [For California Residents](#) | [Terms and Conditions](#) Comments? We would like to hear from you. E-mail us at [customercare@copyright.com](mailto:customercare@copyright.com)



# Evaluating and elucidating the reactivity of OH radicals with atmospheric organic pollutants: Reaction kinetics and mechanisms by machine learning

Flávio O. Sanches-Neto<sup>a,\*</sup>, Jefferson R. Dias-Silva<sup>b</sup>, Vitor M. de Oliveira<sup>b</sup>,  
Vincenzo Aquilanti<sup>c,d</sup>, Valter H. Carvalho-Silva<sup>a,e,\*\*</sup>

<sup>a</sup> Instituto de Química, Universidade de Brasília, Caixa Postal 4478, 70904-970, Brasília, Brazil

<sup>b</sup> Instituto de Química, Universidade Federal de Goiás, Caixa Postal 131, 74001-970, Goiânia-GO, Brazil

<sup>c</sup> Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, 06123, Perugia, Italy

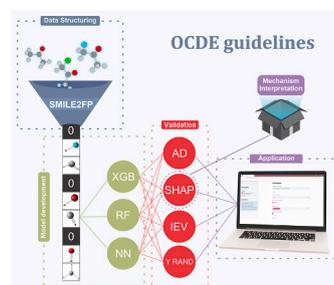
<sup>d</sup> Istituto di Struttura Della Materia, Consiglio Nazionale Delle Ricerche, 00133, Rome, Italy

<sup>e</sup> Modeling of Physical and Chemical Transformations Division, Theoretical and Structural Chemistry Group, Goiás State University, 75132-903, Anápolis, Brazil

## HIGHLIGHTS

- Six machine learning models were developed.
- The half-life time and POCP of the reaction is reported for OH radical.
- Workflow was fitted with experimental data to train ML models using of AOPs.
- The models development followed the OECD principles.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Keywords:

QSAR/QSPR  
OECD guidelines  
Reaction rate constants  
SHAP  
Web application

## ABSTRACT

The rate constants of the reactions of OH radicals with atmospheric organic pollutants (AOPs) are crucial physicochemical parameters to guide in the elucidation of the kinetics and mechanisms of the reactive landscape. The experimental and theoretical difficulties in revealing the reactivity of these degradation processes motivated us to develop a protocol based on machine learning combining molecular fingerprints to estimate their rate constants. The present workflow is based on Organization for Economic Cooperation and Development (OECD) guidelines and state-of-the-art techniques involving (i) data collection including 903 AOPs cataloged in the literature, (ii) pre-processing and structuring of data, (iii) development of models based on three machine learning algorithms, (iv) the standard reference of validation, and (v) mechanistic interpretation. The results show that the built model has a high predictive capacity –  $R_{cv}^2 > 0.959$  and  $RMSE_{cv} < 0.090$  for the training set and  $R_{ext}^2$  and  $Q_{ext}^2 > 0.889$  and  $RMSE_{ext} < 0.084$  for the test set. Additionally, through the SHapley Additive exPlanations (SHAP) method, it was possible to establish insight into the contribution of chemical classes to reaction kinetics and mechanism and to discuss it consistently with current experimental and theoretical observations. The availability of the evaluated reaction rate constants permitted to elucidate the role of AOPs in the photochemical ozone balance. Finally, to disseminate use of our results, we have presented them in a user-

\* Corresponding author.

\*\* Corresponding author. Instituto de Química, Universidade de Brasília, Caixa Postal 4478, 70904-970, Brasília, Brazil.

E-mail addresses: [flavio\\_olimpio@outlook.com](mailto:flavio_olimpio@outlook.com) (F.O. Sanches-Neto), [fatiolog@gmail.com](mailto:fatiolog@gmail.com) (V.H. Carvalho-Silva).

<https://doi.org/10.1016/j.atmosenv.2022.119019>

Received 2 September 2021; Received in revised form 15 February 2022; Accepted 17 February 2022

Available online 25 February 2022

1352-2310/© 2022 Elsevier Ltd. All rights reserved.

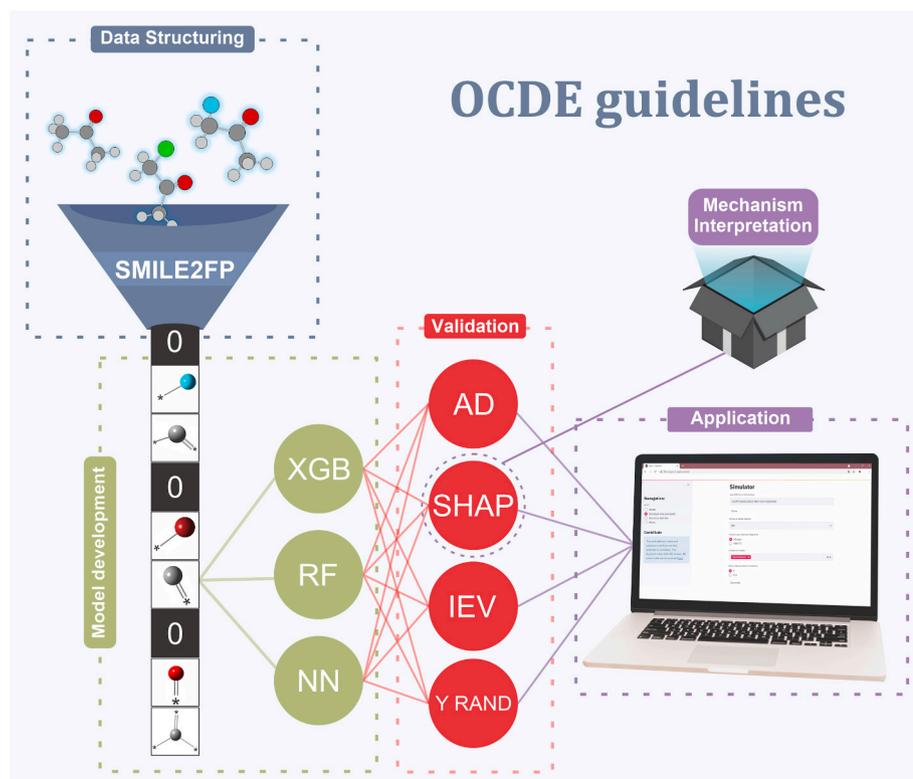
friendly web application that permits compilation of kinetic parameters, and that permits future implementations to account for the temperature dependence in the environmental relevant range, and the consideration of a wider class of chemicals and processes in the mechanistic networks of atmospheric reactivity.

## 1. Introduction

The presence of organic pollutants in the atmosphere has been studied by several research groups around the world (Gupta et al., 2016; Liu et al., 2020). Current and future risks associated with the persistence of these compounds in the atmosphere require specific attention (Nizzetto et al., 2010). Regarding the persistence of these compounds, relevant are compilations reporting data on the mechanism and kinetics of their degradations, that in the troposphere (Tomas, 2005) mainly involve oxidation reactions with OH radicals (Allison, 2016). Hydroxyl radical reactions were investigated by our group in several theoretical works for four-atoms systems, where we extended the understanding of the reactive dynamics accounting for the contribution of quantum tunneling, stereodynamics and roaming (Coutinho et al., 2015). To expand the understanding of the behavior and destination of compounds with more than four atoms, it is important to estimate the reaction rate constant  $k_{OH}$ , which is the most important parameter to reveal the degradation efficiency of a compound (Sudhakaran and Amy, 2013a). However, current protocols providing  $k_{OH}$  are confronted with formidable challenges (Finlayson-Pitts and Pitts Jr, 1999). From an experimental point of view, Atkinson and co-works (Atkinson, 1986; Atkinson and Aschmann, 1984) played an important role in experimental studies to determine the values of  $k_{OH}$  of atmospheric organic pollutants. However, experimental equipment is expensive and system-dependent (Hodson, 1988). Theoretical procedures based on the Transition State Theory combined with calculations of electronic structure have been used to reveal the kinetics and mechanism of degradation of organic pollutants with hydroxyl radicals (Li et al., 2014). On the other hand, theoretical approaches rely on highly expensive computations of potential energy surfaces (Pan et al., 2021; Sanches-Neto et al., 2021a).

Based on experimental and theoretical  $k_{OH}$  data, quantitative structure-activity/property relationships (QSAR/QSPR) models were performed to estimate the rate constant of a set of chemical reactions with the OH radical (Gupta et al., 2016; Öberg, 2005; Wang et al., 2009). To develop a QSAR/QSPR model, specific chemical and computational efforts are needed to select the appropriate molecular descriptors, which are calculated by electronic structures methods and require a high computational cost (Zhong et al., 2020). Additionally, it is necessary to follow the validation principles for regulatory purposes of these models (Gramatica et al., 2004; Öberg, 2005; Sanches-Neto et al., 2021b). Therefore, protocols that do not require experimental approaches and quantum chemistry, both for calculations of molecular descriptors and for estimating the  $k_{OH}$ , are considered relevant. In this sense, the development of ML models combined with molecular fingerprints (MF) has become a successful protocol, since the fingerprint representation procedure is quite simple, *i.e.*, it involves the transformation of a molecular representation into a sequence of binary digits (bits) in order to account for the presence or absence of molecular fragments.

Here, a consolidated accurate machine learning (ML) protocol, according to OECD guidelines (see Fig. 1), is combined with molecular fingerprints to predict  $k_{OH}$  of OH radical with atmospheric organic pollutants (AOPs), quantifying role in the photochemical ozone balance and atmospheric implications based on SHAP (SHapley Additive Planations), a method to interpret ML models (Lundberg et al., 2020).



**Fig. 1.** The routine follows guidelines of the Organization for Economic Cooperation and Development (OECD), see text. Data from the literature, structured and with molecular names converted by SMILES provide molecular fingerprints (MF) – these are the independent variables, while rate constants are the dependent variables of our model developed implemented in three machine learning algorithms – XGboost (XBG), Random Forest (RF), and Neural Network (NN). Validation of the model's performance involved: applicability domain (AD), mechanistic interpretation with the SHAP method, internal and external validation (IEV), and randomization of the dependent variable (Y RAND). Final step is a user-friendly web application.

## 2. Material and methods

### 2.1. Data sets and molecular fingerprint (MF)

The dataset of  $k_{OH}$  with 903 atmospheric pollutants was cataloged from Ref. (Allison, 2016; Wang et al., 2009). Experimental values of the  $k_{OH}$  were collected under standard conditions, 298 K, 101.3 kPa and reported in  $\text{cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ . From the data collected from the literature, a set of steps were carried out to prepare the data for the development of the model. First,  $k_{OH}$  was transformed into a log unit and then scaled from 0 to 1 (using the *sklearn.preprocessing.MinMaxScaler* class in the Python language) to make the dependent variable data with a normal distribution which the consistency was confirmed by Shakiro-Wilk test. Then the compound name or CAS number was converted to SMILES (Simplified Molecular Input Line Entry System) using CIRpy (<https://github.com/mcs07/CIRpy>). Finally, these SMILES were converted into two types of molecular representation with the RDKit program (<https://www.rdkit.org>): (1) MACCS fingerprints and (2) Morgan fingerprints with different bit lengths (512, 1024, 2048, 3072, 4096 and 8192). The full database is available in the file named "SupInfoDataSet.xlsx" in the Supplementary Information file.

### 2.2. Machine learning models

Three ML algorithms - Neural Network (NN) (Hansen and Salamon, 1990), Random Forest (RF) (Ho, 1995) and XGBoost (Chen and Guestrin, 2016) - were built to predict the  $k_{OH}$  of attack of hydroxyl ( $k_{OH}$ ). Random Forest (RF) is one of the most used algorithms in regression and classification models (Hu et al., 2018; Nitze et al., 2012; You et al., 2017). RF is an ensemble method based on a set of decision trees, with each tree having a collection of random variables. The RF algorithm employs randomness when developing the tree architecture, which results in a great diversity producing a better model when compared to other decision tree models (Welbl, 2014; Xing et al., 2019; You et al., 2017).

Another algorithm used to predict the  $k_{OH}$  of atmospheric organic pollutants was eXtreme Gradient Boosting (XGBoost), a method based on a gradient-boosting decision tree (Chen and Guestrin, 2016). This algorithm was developed under the same Gradient Boosting framework and to be highly efficient, flexible, and portable. XGBoost provides a parallel tree reinforcement that solves many data science problems quickly and accurately and has therefore been widely used in recent literature (Li and Zhang, 2018; Meng et al., 2021; Sun et al., 2018; Torlay et al., 2017).

Finally, the last algorithm used in this work was the neural network. This algorithm is based on a series of units organized and connected in sequential layers (Wenzel et al., 2019; Yang et al., 2020). Neural network architecture involves an input layer, hidden layers and an output layer. The units are the neurons, with neurons within the same layer acting in parallel and transforming the input values received from the previous layer into a scalar value. The models were developed using the scikit-learn packages (Géron, 2019; Pedregosa et al., 2011). Details of the hyperparameters of each machine learning model used in this work are listed in Table S2.

### 2.3. Validation

Internal and external validations were necessary to assess the reliability of ML models and to verify their robustness and predictive capacity (Gramatica and Sangion, 2016; Sudhakaran and Amy, 2013b).  $k_{OH}$  data set was randomly split into a training set (80%) and a test set (20%). Performance indices - root-mean-square deviation (RMSE), correlation coefficient ( $R^2$ ), Pearson correlation coefficient of prediction ( $r_{ext}^2$ ), predictive power ( $Q_{cv}^2$ ), and external validation ( $Q_{ext}^2$ ) - were calculated using the formulas:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{n}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y}_{exp})^2} \quad (2)$$

$$r_{ext}^2 = \frac{\sum_{i=1}^n (y_{exp} - \bar{y}_{exp}) \cdot (y_{pred} - \bar{y}_{pred})}{\sqrt{\sum_{i=1}^n (y_{exp} - \bar{y}_{exp})^2} \sqrt{\sum_{i=1}^n (y_{pred} - \bar{y}_{pred})^2}} \quad (3)$$

$$Q_{cv}^2 = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y}_{exp})^2} \quad (4)$$

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^n (y_{exp} - y_{pred})^2}{\sum_{i=1}^n (y_{exp} - \bar{y}_{exp}^r)^2} \quad (5)$$

where  $y_{exp}$ ,  $y_{pred}$ ,  $\bar{y}_{exp}$ , and  $\bar{y}_{pred}$  are the experimental, predicted, average of the experimental and predicted values of the dependent variable (over the experimental and predicted values of the dependent variable (over the validation set), respectively.  $\bar{y}_{exp}^r$  is the average value of the dependent variable for the training set - the sums cover all the compounds in the validation set. The subscript *cv* and *ext* represent the data obtained from the internal (cross-validation) and external validation datasets, respectively. For the training data set, a 10-fold cross-validation method was employed (using the *sklearn.model\_selection.StratifiedKFold()* class in the Python language), which randomly divided the data into 10 subsets. With subsequent optimization cycles, one subset was retained for validation and the others were used for training, this procedure was repeated 10 times. External validation (test data set) for ML models was applied, assuming a subgroup excluded from the training set. Furthermore, the *p*-value and residual plots were also studied as a validation measure. The *p*-value indicates the conditional probability that a relationship as strong as that observed in the data would be present, if the null hypothesis is true (Sudhakaran and Amy, 2013a). Consequently, a low *p*-value (<0.05) is required to build a statistically significant ML model. If the points on a residual plot are randomly scattered around zero on the horizontal axis, a linear regression model is appropriate for the data.

### 2.4. Applicability domain

To assess whether the models developed in this workflow have a reliable prediction, the applicability domain (AD), crucial for any QSAR/QSPR model (Gramatica, 2014; Netzeva et al., 2005; OECD, 2007), was calculated. The AD aims to compare the similarity of the query compound with the compounds in the training data set. If there is a significant similarity, defined by a predefined threshold, the query compound will be reliably predicted by our models. The Tanimoto index (Tanimoto, 1958)  $T_c(A, B)$  was used to evaluate the similarity between two compounds, A and B, for both types of MFs and radicals, according to the following equation:

$$T_c(A, B) = \frac{c}{a + b - c} \quad (6)$$

where *a* and *b* are the number of structural features for compound A and B, respectively, or bits set to 1, in each molecule, *c* is the number in common.

The maximum similarity, which refers to the maximum value of the Tanimoto index between all similarity values obtained, and the mean similarity, which refers to the mean of these similarity values, were used as metrics to assess whether the query compound is inside or outside the applicability domain. For each predefined threshold, compounds that were outside the applicability domain were removed from the test dataset and  $RMSE_{ext}$  was recalculated. The optimal limit was the one

that obtained the lowest  $RMSE_{ext}$  value with the lowest possible number of compounds outside the applicability domain.

### 2.5. Model interpretation

Another fundamental requirement in developing a QSAR model is to understand how the model is performed a prediction – one of the most important validation principles in the development of models, whenever possible to be carried out, is the mechanistic interpretation. (OECD, 2007). In this sense, a recent unified approach to interpreting model predictions, SHAP (Lundberg et al., 2020a), which sheds light on the black box of ML algorithms, was developed to elucidate the most important features learned by the model (Zhong et al., 2021). The SHAP approach was derived from cooperative game theory – primarily developed to estimate the importance of each player on a team (Lundberg et al., 2018). For this, a reward for each player is carried out depending on their importance and their contributions to the result of a game. For our case, the use of molecular fingerprints, the Shapley values provide a solution for assigning a fair or reasonable reward for each of the Morgan and MACCS fingerprint characteristics and represent a unique result. The following equation is used to calculate the Shapley value  $\Phi_i$ :

$$\Phi_i = \frac{1}{|N|!} \sum_{S \subseteq N} |S|!(|N| - |S| - 1)! [f(S \cup \{i\}) - f(S)]$$

where  $f(S)$  corresponds to the output of the ML model to be explained using a set  $S$  of features, and  $N$  is the complete set of all features. The final contribution or Shapley value of feature  $i$  ( $\Phi_i$ ) is determined as the average of its contributions across all possible permutations of a feature set. The predictions for all possible subsets  $S \subseteq N$  are calculated because the effect of withholding a feature depends on all other features in the model.

## 3. Results and discussion

### 3.1. Internal and external validation

In this work, were applied as detailed in Fig. 1: data collection, data preprocessing, model development, model validation and interpretation – all protocols for the development of QSAR models according to the Organization for Economic Cooperation and Development (OECD, 2007).

From a set of data cataloged of the literature (Allison, 2016; Wang et al., 2009), a generic workflow with three ML algorithms (Chen and Guestrin, 2016; Ho, 1995) combined with two types of molecular fingerprints of the AOPs was adjusted to predict the reactivity of these compounds under atmospheric conditions. The data were randomly divided into a training set (80%) and a test set (20%) and the main statistical criteria were performed for internal and external validation of the developed models (Gramatica, 2014). Fig. S1 shows the effect of MF length for the Morgan fingerprint and a more detailed discussion is provided in SI.

The validation of the six models developed was performed both for the training set, internal validation, as for the test set, external validation and the values of the statistical parameters are listed in Table 1. The statistical criteria  $R_{cv}^2$ ,  $Q_{cv}^2$  and  $RMSE_{cv}$  were calculated to evaluate the performance of the models using the 10-fold cross-validation method (see sec 2.3) in the internal validation. Similarly, for external validation,  $r_{ext}^2$ ,  $R_{ext}^2$ ,  $Q_{ext}^2$ , and  $RMSE_{ext}$  were selected to measure the quality of the models. The results show that the built model has a high predictive capacity –  $R_{cv}^2 > 0.959$  and  $RMSE_{cv} < 0.090$  for the training set and  $R_{ext}^2$  and  $Q_{ext}^2 > 0.889$  and  $RMSE_{ext} < 0.084$  for the test set. All these results provided a good predictive capacity and clearly indicated that our models exclude overfitted behavior. The difference between the  $RMSE_{ext}$  and  $Q_{ext}^2$  values of the test set were used to assess the performance of ML

**Table 1**

Internal and external validation parameters of the ML models applied to  $k_{OH}$  data set with Morgan (4096 length of molecular fingerprint) and MACCS keys fingerprints.

Models	Morgan Fingerprint					
	Training set			Test set		
	$R_{cv}^2$	$Q_{cv}^2$	$RMSE_{cv}$	$r_{ext}^2$	$R_{ext}^2   Q_{ext}^2$ <sup>a</sup>	$RMSE_{ext}$
NN	0.993	0.752	0.084	0.922	0.843	0.073
XGB	0.972	0.713	0.090	0.905	0.817	0.079
RF	0.960	0.723	0.088	0.889	0.789	0.084
Models	MACCS Fingerprint					
	Training set			Test set		
	$R_{cv}^2$	$Q_{cv}^2$	$RMSE_{cv}$	$r_{ext}^2$	$R_{ext}^2   Q_{ext}^2$	$RMSE_{ext}$
NN	0.973	0.802	0.074	0.950	0.888	0.062
XGB	0.974	0.789	0.077	0.960	0.921	0.052
RF	0.959	0.763	0.082	0.946	0.896	0.059

<sup>a</sup> The values are equal to the third decimal number.

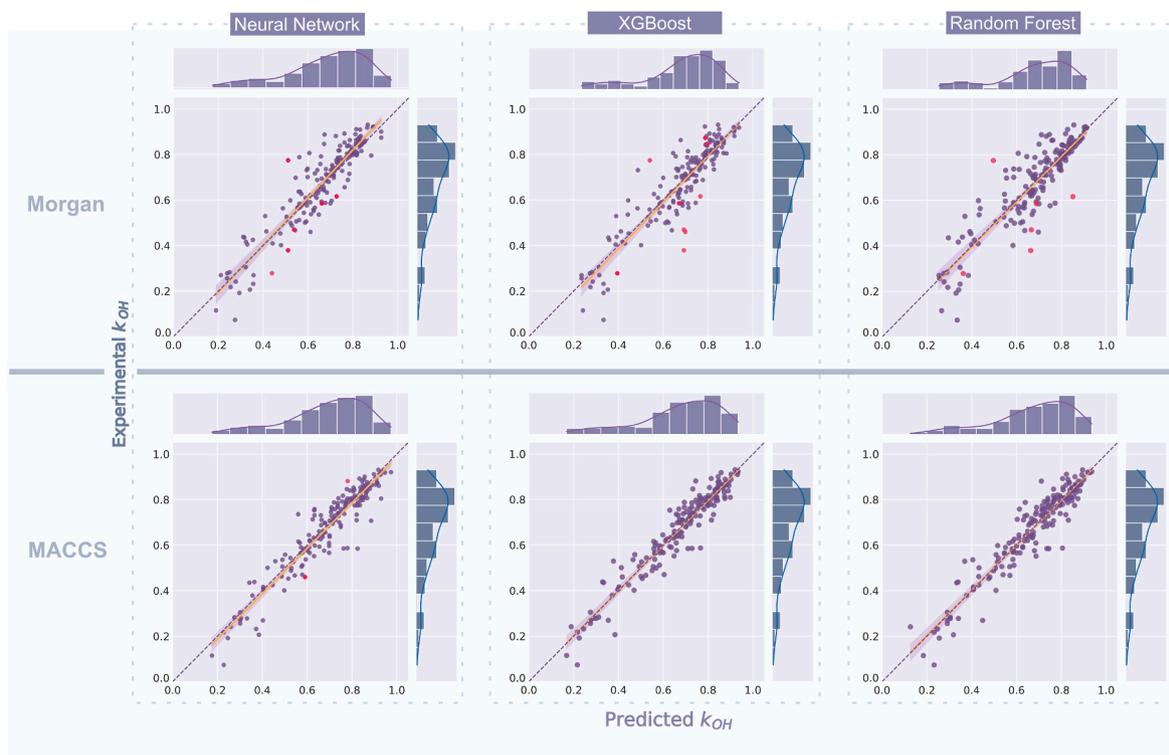
models to predict the  $k_{OH}$ . Although all models have values similar to  $RMSE_{ext}$  and  $Q_{ext}^2$ , the XGB algorithm combined with MACCS fingerprint was the best performing model and yield results highly suitable for both the training set and the test set.

### 3.2. Predicted vs experimental data

Fig. 2 shows the predictions and distributions of the data obtained by the six models in a comparison chart on the experimental data. The purple dashed line is the reference line ( $r_{ext}^2 = 1$ ) and indicates predictive on capacity of the ML models – deviations from the reference line indicating that the data are under or overestimated closeness to the reference line indicating good-prediction *i.e.*, on the other hand. Name of compounds and their coordinates, x (predicted) and y (experimental) axis, in Fig. 2 represented in red are listed in Table S1. Additionally, all models showed a  $p$ -value  $< 0.05$ , indicating that for the 95% confidence interval,  $r_{ext}^2$  is statistically significant. Fig. S2 (see SI file) shows the residual plots all models studied and as expected, the residues are randomly distributed around zero. All these results provided a good predictive capacity and clearly indicated that our models excluded overfitted behavior. To further confirm the quality of the data, a robustness test was performed by applying the Y randomization technique (Fan et al., 2018; Rücker et al., 2007), to exclude the possibility of casual correlation. This procedure indicated that all new values calculated from  $Q_{ext}^2$  were much lower than the original model, suggesting that the results of the original model were not accidental (see Table S3).

### 3.3. Applicability domain

To verify the chemical domain in which our model is useful to be applied, the applicability domain of all the models developed in this work were determined by means of the Tanimoto index (Tanimoto, 1958). Table 2 shows the optimal values found for the AD for the two types of MF – Morgan and MACCS. The threshold of 0.290 led to the lowest  $RMSE_{ext}$  value for the NN and RF models and the limit of 0.3100 for the XGB model using Morgan Fingerprint. For the MACCS Fingerprint, the limit of 0.500 yield the lowest  $RMSE_{ext}$  values for the NN and RF models and the limit of 0.530 for the XGB. The maximum similarity metric was chosen to indicate whether the compound is within or outside the applicability domain, *i.e.*, if the query compound has a value less than the maximum similarity threshold value, the compound will be outside the AD, otherwise, the compound will be within the AD of the model developed. The structure of the outliers in this study is shown in Fig. S3. The MACCS MF yielded smaller deviations based on model predictions when compared to the Morgan MF. For example, Morgan fingerprints decrease their predictive capacity with the presence of



**Fig. 2.** Plot of predicted vs experimental values for data sets of the oxidative process mediated by OH radicals in atmospheric organic pollutants. External validation of the ML models is performed comparing the reference line (dashed) with the linear representation of the data (the distribution of the  $k_{OH}$  values is shown. Experimental (right in purple) and predicted (top in blue). Points highlighted in red are considered outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 2**

The thresholds of similarity, the number of compounds outside the AD for each threshold value, and the corresponding  $RMSE_{ext}$ .

Morgan Fingerprint								
NN			XGB			RF		
Threshold	$RMSE_{ext}$	NCOAD <sup>a</sup>	Threshold	$RMSE_{ext}$	NCOAD	Threshold	$RMSE_{ext}$	NCOAD
0.290	0.0688	6	0.310	0.0686	9	0.290	0.0765	6
MACCS Fingerprint								
NN			XGB			RF		
Threshold	$RMSE_{ext}$	NCOAD	Threshold	$RMSE_{ext}$	NCOAD	Threshold	$RMSE_{ext}$	NCOAD
0.530	0.0602	2	0.500	0.0517	0	0.500	0.0594	0

a Number of compounds outside the AD.

isotopes, such as Propene-d6. The presence of isotopes is quite important, as the kinetic isotopic effect plays an important role in atmospheric chemistry (Anderson et al., 2004; Iannone et al., 2003). Therefore, a larger amount of  $k_{OH}$  data involving isotopic variants is relevant for a larger applicability domain of our model. Furthermore, according to the model predictions, the higher relative error deviations are mostly typical of small-size molecules and groups such as amines (N) and halogen (especially, F and I).

### 3.4. Comparison with previous studies

Most studies for the prediction of  $k_{OH}$  were used with linear, multiple linear regression (MLR) and partial least squares (PLS) (Gramatica et al., 2004; Öberg, 2005; Roy et al., 2011; Wang et al., 2009), ensembles, decision tree forest (DTF) and decision tree boost (DTB) (Gupta et al., 2016) and nonlinear (NN) (Allison, 2016) methods. Another important factor investigated in these studies was the effect of temperature on  $k_{OH}$ , and only two studies appear to have been reported on temperature dependence. The work of Li and Gupta regarded the study of the

prediction of  $k_{OH}$  in the temperature range from 206 to 1364 K, while the other works were carried out at room temperature (298 K).

To the best of our knowledge, this is the first study that employs ML combined with molecular fingerprint to estimate the  $k_{OH}$  of AOPs with OH radical. Our work employs two ensemble methods as performed by Gupta and a nonlinear method, neural networks, as employed by Allison; this last work is not compared to in this study as it does not employ the conventional metrics of the other studies. A comparison with the performance indices of the models is shown in Table 3. As can be seen, the present work presents the largest amount of data for a single temperature (298 K), without the use of molecular descriptors derived from quantum chemistry calculations, and the statistical criteria are either similar or even superior to previous studies.

### 3.5. Interpretation and insight

Besides statistical performances of each model, it is particularly important to know its physicochemical and mechanistic predictive insight (Zhong et al., 2021), elucidating ML model features versus

**Table 3**Comparison with previous works between different models for internal (training set) and external (test set) validation in the prediction of  $k_{OH}$  values.

Model	Algorithm	n <sup>a</sup>	p <sup>b</sup>	t <sup>c</sup> (Kelvin)	Training Set		Test Set	
					R <sup>2</sup> <sub>cv</sub>	Q <sup>2</sup> <sub>cv</sub>	R <sup>2</sup> <sub>ext</sub>	Q <sup>2</sup> <sub>ext</sub>
Gramatica (2004a)	MLR <sup>d</sup>	460	4	298	0.828	0.819	–	0.826
Gramatica (2004a)	MLR	460	4	298	0.828	0.816	–	0.813
Gramatica (2004b)	MLR	460	6	298	0.846	0.841	–	0.866
Öberg (2005)	PLS <sup>e</sup>	743	333	298	0.906	0.875	–	0.840
Wang et al. (2009)	PLS	722	22	298	0.878	0.865	–	0.872
Roy et al. (2011)	MLR	460	4	298	0.824	0.819	–	–
Li et al. (2014)	MLR	872	12	298	0.883	0.879	0.858	0.851
Li et al. (2014)	MLR	1543	14	206–1364	0.873	0.871	0.838	0.835
Gupta et al. (2016)	DTF	1543	4	206–1364	0.900	–	0.910	–
Gupta et al. (2016)	DTB	1543	4	206–1364	0.920	–	0.920	–
Liu et al. (2020)	MLR	180	5	298	0.785	0.754	–	0.642
This work	XGB-MACCS	903	1	298	0.974	0.792	0.896	0.896

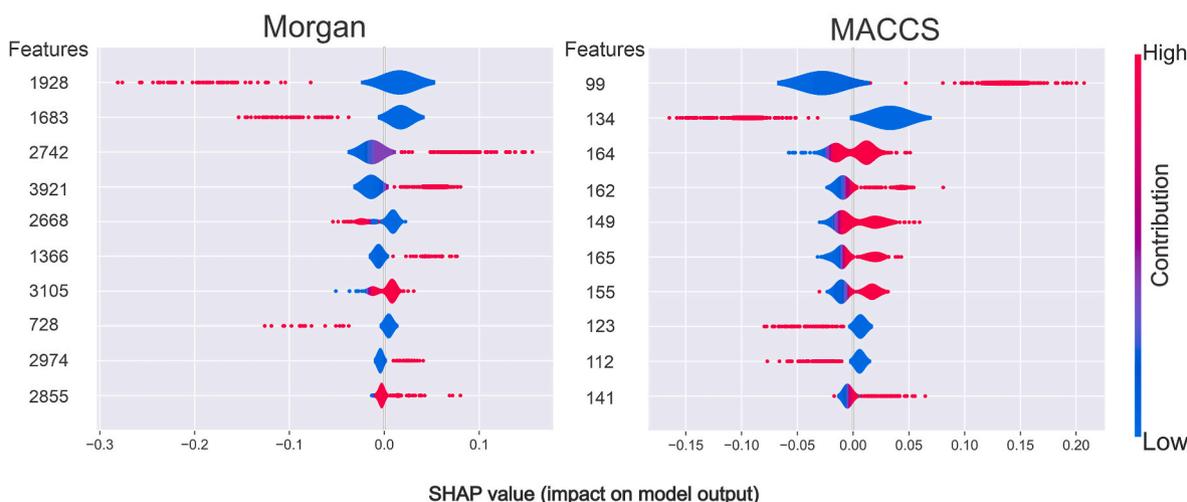
<sup>a</sup> n = the total number of chemical compounds in the dataset.<sup>b</sup> p = the number of molecular descriptors.<sup>c</sup> Temperature range.<sup>d</sup> MLR: multiple linear regression.<sup>e</sup> PLS: partial least squares.

expert, experimental techniques and sophisticated theories (Lundberg et al., 2020b). A recent unified approach to interpret the predictions of the model, SHAP (SHapley Additive exPlanations) brings light into the “black box” of ML algorithms, was employed to confront the top 10 learned characteristics learned by the XGBoost model with the two molecular fingerprints developed in this work (Fig. 3). The blue and red colors denote absence or presence of a given molecular fragment in a compound, and the SHAP value (x-axis) shows the impact that fragment has on the prediction. For example, features 2742 (Morgan) and 99 (MACCS), involving the presence of a C=C bond, increase the reactivity of a given compound, consistently with experimental observations (Atkinson et al., 1982; Atkinson and Aschmann, 1984). Similarly, features 1928, 1683 and 3105 in Morgan involve the presence of F, Cl and Br, respectively, while feature 134 represents a halogen in MACCS: the presence of this feature in organic pollutants correlates with a decreased reactivity, also consistently with literature (Atkinson, 1986; Li et al., 2014). Tables S4 and S5 of SI identifies fragments regarding both fingerprints and specific role on reactivity.

In order to obtain a more in-depth interpretation of the model, the highest SHAP values of a series of compounds were analyzed to verify whether the model learned the reactive sites in organic pollutants through the reaction with the OH radical. For each chemical class, the SHAP plot, the chemical structure and the most important fragments of

the molecules are presented in Fig. 4. The length of the bar indicates the SHAP value of each fragment, and the red or blue colors indicate whether it contributes to increase or to decrease the reactivity.

Regarding alkanes, it is known that OH radicals react via abstraction of the hydrogen atom of the C–H bond (Atkinson, 1986), here consistently shown by SHAP, feature 149 (CH<sub>3</sub> group). Regarding alkenes, the mechanism proceeds via addition to the double bond, clearly indicating that the model has correctly learned preference of a double bond (feature 99) over a simple bond. When a halogen is added in an alkane or alkene, forming haloalkanes and haloalkenes, the mechanism occurs through H atom abstraction and addition to the double bond, respectively. However, the presence of a halogen, an electron-withdrawing group, provides a decreased reactivity, again correctly identified by SHAP (feature 134). Esters and carboxylic acids are compounds that play an important role in the troposphere and their degradation mechanisms also occur through hydrogen abstraction (Ren et al., 2019). The SHAP revealed that methyl group (feature 149), where abstractions occur, are characteristics that increase reactivity and that the OCO group (feature 123) is associated with a negative effect on reactivity, corroborating the experimental information. The reactions of the OH radical with ketones and aldehyde are considered an atmospheric sinkhole due to the abundance of these compounds and their mechanism – which also occurs through the abstraction of the H atom (Ponnusamy



**Fig. 3.** Importance of the top ten representative molecular fingerprints and the SHAP values for the ML models applied to  $k_{OH}$  data set and based on Morgan and MACCS fingerprints. See text and SI.

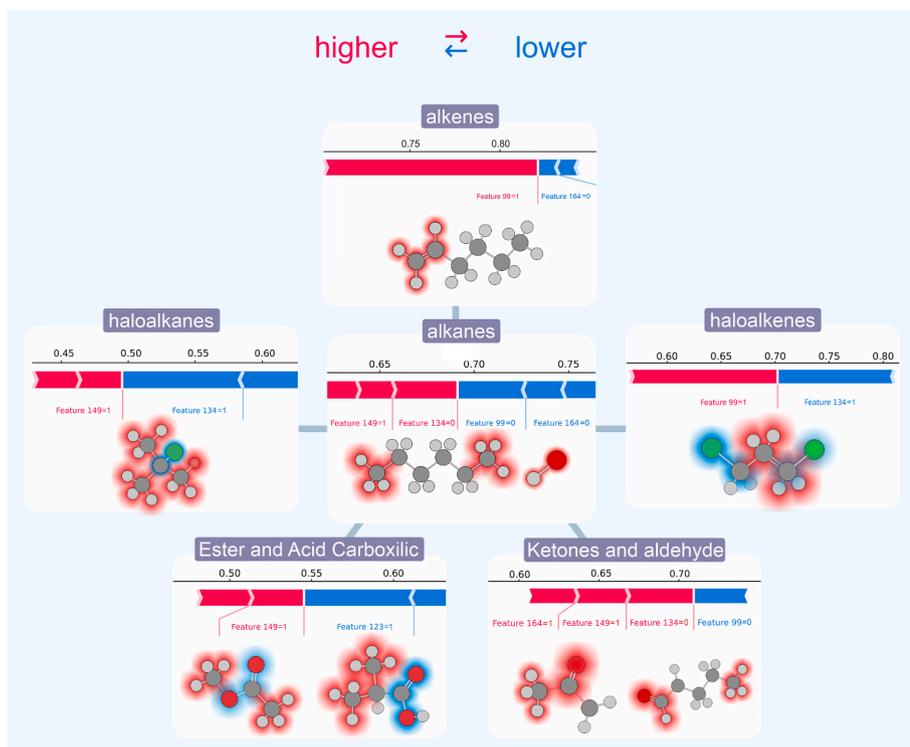


Fig. 4. Pictorial representation of the force plot exhibiting the largest SHAP values for different chemical classes learned by the XGB-MACCS model. The length of each bar represents the SHAP value for a given feature – the longer the bar, the higher the SHAP value is. The SHAP values rank from the highest at the boundary between the two colors to the smallest on either side for both red and blue bars. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2018). Although the presence of the carbonyl group has no effect on reactivity, the SHAP values show that the presence of the oxygen atom (feature 164) contributes positively to the reactivity like the  $\text{CH}_3$  group (feature 149), according to evidence from kinetic data.

All this heuristic protocol reveals that our model has learned the essential characteristics of the molecular system, not spurious, clearly indicating that it does not have a compensating effect between precision and interpretability: a theoretical algorithm that does not balance the quality of statistical parameterization and physical-chemical interpretation must be avoided, as it can lead to distorted views of fundamental chemical processes running away from chemical reality and creating *ad hoc* interpretations. An additional advantage of our approach is that no need of molecular descriptor calculations is required to predict the  $k_{\text{OH}}$ .

### 3.6. Kinetic parameters and web application

From the data of the  $k_{\text{OH}}$  for removal of organic pollutants with the OH radical, as obtained by ML models, it is immediate to obtain pseudo-unimolecular half-life times using  $t_{1/2} = \ln 2 / (k_{\text{OH}} \times [\text{OH}])$ , where  $[\text{OH}]$  is the average concentration of OH radicals in the troposphere. The half-life time of the reaction is reported at the temperature of 298.15 K, and  $[\text{OH}]$   $5.0 \times 10^5$ – $5.0 \times 10^6$  molecules  $\text{cm}^{-3}$ , which usually represents the values found in troposphere (Öberg, 2005; Prinn et al., 1995). To calculate the ability of AOP's to create ozone in the atmosphere through the reaction initiated by OH radical, the photochemical ozone creation potential (POCP) (Jenkin and Hayman, 1999) was calculated using the following expression:

$$\varepsilon_{\text{POCP}} = \alpha_1 \cdot \gamma_s \cdot \gamma_R^\beta \cdot (1 - \alpha_2 \cdot n_c) \quad (7)$$

where  $\varepsilon_{\text{POCP}}$  is the estimated POCP –  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$  are constants, with the values of 111, 0.04 and 0.5, respectively.  $\gamma_s$  is the structure-based ozone formation index,  $\gamma_R$  is the reactivity-based ozone formation index, and  $n_c$  is the carbon numbers of the compound. The structure and reactivity-based ozone formation indices are further defined as:

$$\gamma_s = \frac{n_B}{MW} \frac{28.05}{6} \quad (8)$$

$$\gamma_R = \frac{k_{\text{OH}}}{n_B} \frac{6}{k_{\text{OH}}^{\text{ethene}}} \quad (9)$$

where  $n_B$  is the total number of C–C and C–H bonds in the molecule, MW is the molecular weight,  $k_{\text{OH}}$  is the rate constant for reaction with OH radicals at 298 K and 101.3 kPa of air, and  $k_{\text{OH}}^{\text{ethene}}$  is the rate constant for reaction of ethene with OH radicals at 298 K and 101.3 kPa of air ( $8.64 \times 10^{-12} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ ) (Derwent et al., 1998; Jenkin and Hayman, 1999).

Finally, to share our results, we have added our model in a user-friendly web application to make predictions of the  $k_{\text{OH}}$ , half-life and POCP of the different organic pollutants with the OH radical, which is available in [www.pysirc.com.br](http://www.pysirc.com.br). To extract the predictions, the only information required from the user are the SMILES or CAS number of the molecule.

## 4. Conclusion and remarks

A protocol is presented based on machine learning combined with molecular fingerprints providing  $k_{\text{OH}}$  and reaction half-life of OH radical with organic pollutants in atmospheric environments. Three algorithms of ML combined with two types of MF were developed. To evaluate the performance of the built models, the main statistical criteria were calculated and yielded excellent predictive capacity. Regarding the recently published works, our results show higher values for the statistical criteria of the external validation set, suggesting that our models have greater predictive capacity to estimate  $k_{\text{OH}}$  with no need of molecular descriptors. Finally, an interpretation of the models, performed using the SHAP, indicated that the predictions of the models were based on a reasonable understanding, specifically of how electron-withdrawing and -donating groups interfere of the reactivity of the OH radical.

To disseminate use of our results, we have presented them in a user-

friendly web application that permits compilation of kinetic parameters: future progress of this investigation will include (i) important explicit account of the temperature dependence in the environmental relevant range, having recently being demonstrated that often deviate can occur from the usual Arrhenius behavior (Aquilanti et al., 2017; Carvalho-Silva et al., 2019), (ii) consideration along these lines of a richer class of chemical compounds and processes in models of reactive mechanistic networks (Eisenreich et al., 2021), and (iii) insertion within currently operating chemical kinetic codes (Machado et al., 2019), in particular implementing progress in understanding reactivity in extreme environments, such as those relevant in planetary sciences and astrochemistry.

### CRedit authorship contribution statement

**Flávio O. Sanches-Neto:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Investigation, Data curation. **Jefferson R. Dias-Silva:** Methodology, Formal analysis, Investigation, Validation. **Vitor M. de Oliveira:** Methodology, Formal analysis, Investigation, Validation. **Vincenzo Aquilanti:** Writing – review & editing, Visualization, Formal analysis, Supervision. **Valter H. Carvalho-Silva:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The authors acknowledge the Brazilian funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support. Valter H. Carvalho-Silva thanks Goiana agency FAPEG for the research funding programs (AUXÍLIO À PESQUISA COLABORATIVA FAPEG-FAPESP/GSP2019011000037). Flávio O. Sanches-Neto thanks the University of Brasília (AUXÍLIO FINANCEIRO A ESTUDANTES E PESQUISADORES - Edital DPG N° 0007/2021).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.atmosenv.2022.119019>.

### References

- Allison, T.C., 2016. Application of an artificial neural network to the prediction of OH radical reaction rate constants for evaluating global warming potential. *J. Phys. Chem. B* 120, 1854–1863. <https://doi.org/10.1021/acs.jpcc.5b09558>.
- Anderson, R.S., Iannone, R., Thompson, A.E., Rudolph, J., Huang, L., 2004. Carbon kinetic isotope effects in the gas-phase reactions of aromatic hydrocarbons with the OH radical at 296 ± 4 K. *Geophys. Res. Lett.* 31, 1–4.
- Aquilanti, V., Coutinho, N.D., Carvalho-Silva, V.H., 2017. Kinetics of low-temperature transitions and reaction rate theory from non-equilibrium distributions. *Phil. Trans. Roy. Soc. Lond.* 375, 20160204.
- Atkinson, R., 1986. Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds under atmospheric conditions. *Chem. Rev.* 86, 69–201. <https://doi.org/10.1021/cr00071a004>.
- Atkinson, R., Aschmann, S.M., 1984. Rate constants for the reaction of OH radicals with a series of alkenes and dialkenes at 295 K. *Int. J. Chem. Kinet.* 16, 1175–1186.
- Atkinson, R., Aschmann, S.M., Winer, A.M., Pitts Jr., J.N., 1982. Rate constants for the reaction of OH radicals with a series of alkanes and alkenes at 299 K. *Int. J. Chem. Kinet.* 14, 507–516.
- Carvalho-Silva, V.H., Coutinho, N.D., Aquilanti, V., 2019. Temperature dependence of rate processes beyond Arrhenius and eyring: activation and transitionity. *Front. Chem.* 7, 380. <https://doi.org/10.3389/fchem.2019.00380>.

- Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.
- Coutinho, N.D., Silva, V.H.C., de Oliveira, H.C.B., Camargo, A.J., Mundim, K.C., Aquilanti, V., 2015. Stereodynamical origin of anti-arrhenius kinetics: negative activation energy and roaming for a four-atom reaction. *J. Phys. Chem. Lett.* 6, 1553–1558. <https://doi.org/10.1021/acs.jpcclett.5b00384>.
- Derwent, R.G., Jenkin, M.E., Saunders, S.M., Pilling, M.J., 1998. Photochemical ozone creation potentials for organic compounds in northwest Europe calculated with a master chemical mechanism. *Atmos. Environ.* 32, 2429–2441.
- Eisenreich, S.J., Hornbuckle, K., Jones, K.C., 2021. The global legacy of POPs: special issue. *Environ. Sci. Technol.* 55, 9397–9399. <https://doi.org/10.1021/ACS.EST.1C03067>.
- Fan, D., Yang, H., Li, F., Sun, L., Di, P., Li, W., Tang, Y., Liu, G., 2018. In silico prediction of chemical genotoxicity using machine learning methods and structural alerts. *Toxicol. Res. (Camb)* 7, 211–220.
- Finlayson-Pitts, B.J., Pitts Jr., J.N., 1999. *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications*. Elsevier.
- Géron, A., 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.
- Gramatica, P., 2014. External evaluation of QSAR models, in addition to cross-validation: verification of predictive capability on totally new chemicals. *Mol. Inform.* 33, 311–314. <https://doi.org/10.1002/minf.201400030>.
- Gramatica, P., Pilutti, P., Papa, E., 2004. Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training- test sets and consensus modeling. *J. Chem. Inf. Comput. Sci.* 44, 1794–1802.
- Gramatica, P., Sangion, A., 2016. A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. *J. Chem. Inf. Model.* 56, 1127–1131.
- Gupta, S., Basant, N., Mohan, D., Singh, K.P., 2016. Modeling the reactivities of hydroxyl radical and ozone towards atmospheric organic chemicals using quantitative structure-reactivity relationship approaches. *Environ. Sci. Pollut. Res.* 14034–14046. <https://doi.org/10.1007/s11356-016-6527-2>.
- Hansen, L.K., Salamon, P., 1990. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 993–1001.
- Ho, T.K., 1995. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*, pp. 278–282.
- Hodson, J., 1988. The estimation OF the photodegradation OF organic compounds BY hydroxyl radical reaction rate constants obtained from nuclear magnetic resonance spectroscopy chemical shift data. *Chemosphere* 17, 2339–2348.
- Hu, S., Liu, H., Zhao, W., Shi, T., Hu, Z., Li, Q., Wu, G., 2018. Comparison of machine learning techniques in inferring phytoplankton size classes. *Rem. Sens.* 10, 191.
- Iannone, R., Anderson, R.S., Rudolph, J., Huang, L., Ernst, D., 2003. The carbon kinetic isotope effects of ozone-alkene reactions in the gas-phase and the impact of ozone reactions on the stable carbon isotope ratios of alkenes in the atmosphere. *Geophys. Res. Lett.* 30, 1–4.
- Jenkin, M.E., Hayman, G.D., 1999. Photochemical ozone creation potentials for oxygenated volatile organic compounds: sensitivity to variations in kinetic and mechanistic parameters. *Atmos. Environ.* 33, 1275–1293.
- Li, C., Xie, H., Chen, J., Yang, X., Zhang, Y., Qiao, X., 2014. Predicting gaseous reaction rates of short chain chlorinated para ffn with -OH: overcoming the di ffi culty in experimental determination. *Environ. Sci. Technol.* 48, 13808–13816. <https://doi.org/10.1021/es504339r>.
- Li, P., Zhang, J.-S., 2018. A new hybrid method for China's energy supply security forecasting based on arima and xgboost. *Energies* 11, 1687.
- Liu, Y., Cheng, Z., Liu, S., Tan, Y., Yuan, T., Yu, X., Shen, Z., 2020. Quantitative structure activity relationship (QSAR) modelling of the degradability rate constant of volatile organic compounds (VOCs) by OH radicals in atmosphere. *Sci. Total Environ.* 729, 138871.
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020a. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 2522–2539.
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S., 2020b. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- Lundberg, S.M., Lee, S.-I., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., others, 2018. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., p. 749
- Machado, H.G., Sanches-Neto, F.O., Coutinho, N.D., Mundim, K.C., Palazzetti, F., Carvalho-Silva, V.H., 2019. Transitivity: a code for computing kinetic and related parameters in chemical transformations and transport phenomena. *Molecules* 24, 3478.
- Meng, Y., Yang, N., Qian, Z., Zhang, G., 2021. What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. *J. Theor. Appl. Electron. Commer. Res.* 16, 466–490.
- Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., others, 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Altern. to Lab. Anim.* 33, 155–173.
- Nitze, I., Schulthess, U., Asche, H., 2012. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. 4th GEOBIA* 35.

- Nizzetto, L., Macleod, M., Borgå, K., Cabrerizo, A., Dachs, J., Guardo, A., Di Ghirardello, D., Hansen, K.M., Jarvis, A., Lindroth, A., others, 2010. Past, present, and future controls on levels of persistent organic pollutants in the global environment. *Environ. Sci. Technol.* 44, 6526–6531.
- Öberg, T., 2005. A QSAR for the hydroxyl radical reaction rate constant: validation, domain of application, and prediction. *Atmos. Environ.* 39, 2189–2200.
- Oecd, O., 2007. Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. In: Organisation for Economic Co-operation and Development: Paris, France.
- Pan, W., Chang, J., He, S., Xue, Q., Liu, X., Fu, J., Zhang, A., 2021. Major influence of hydroxyl and nitrate radicals on air pollution by environmentally persistent free radicals. *Environ. Chem. Lett.* 1–7. <https://doi.org/10.1007/s10311-021-01278-9>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., others, 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Ponnusamy, S., Sandhiya, L., Senthikumar, K., 2018. Atmospheric oxidation mechanism and kinetics of hydrofluoroethers, CH<sub>3</sub>OCHF<sub>3</sub>, CH<sub>3</sub>OCHF<sub>2</sub>, and CHF<sub>2</sub>OCH<sub>2</sub>CF<sub>3</sub>, by OH radical: a theoretical study. *J. Phys. Chem.* 122, 4972–4982.
- Prinn, R.G., Weiss, R.F., Miller, B.R., Huang, J., Alyea, F.N., Cunnold, D.M., Fraser, P.J., Hartley, D.E., Simmonds, P.G., 1995. Atmospheric trends and lifetime of CH<sub>3</sub>CCl<sub>3</sub> and global OH concentrations. *Science* (80-) 269, 187–192.
- Ren, Y., Cai, M., Daële, V., Mellouki, A., 2019. Rate coefficients for the reactions of OH radical and ozone with a series of unsaturated esters. *Atmos. Environ.* 200, 243–253.
- Roy, P.P., Kovarich, S., Gramatica, P., 2011. QSAR Model Reproducibility and Applicability : A Case Study of Rate Constants of Hydroxyl Radical Reaction Models Applied to Polybrominated Diphenyl Ethers and (Benzo-) Triazoles. <https://doi.org/10.1002/jcc>.
- Rücker, C., Rücker, G., Meringer, M., 2007.  $\gamma$ -Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* 47, 2345–2357.
- Sanches-Neto, F.O., Ramos, B., Lastre-Acosta, A.M., Teixeira, A.C.S.C., Carvalho-Silva, V. H., 2021a. Aqueous picloram degradation by hydroxyl radicals: unveiling mechanism, kinetics, and ecotoxicity through experimental and theoretical approaches. *Chemosphere* 278, 130401. <https://doi.org/10.1016/j.chemosphere.2021.130401>.
- Sanches-Neto, F.O., Richard Dias-Silva, J., Henrique Keng Queiroz Junior, L., Henrique Carvalho-Silva, V., 2021b. “pySiRC”: machine learning combined with molecular fingerprints to predict the reaction rate constant of the radical-based oxidation processes of aqueous organic contaminants. *Environ. Sci. Technol.* 55, 12437–12448. <https://doi.org/10.1021/acs.est.1c04326>.
- Sudhakaran, S., Amy, G.L., 2013a. QSAR models for oxidation of organic micropollutants in water based on ozone and hydroxyl radical rate constants and their chemical classification. *Water Res.* 47, 1111–1122. <https://doi.org/10.1016/j.watres.2012.11.033>.
- Sudhakaran, S., Amy, G.L., 2013b. QSAR models for oxidation of organic micropollutants in water based on ozone and hydroxyl radical rate constants and their chemical classification. *Water Res.* 47, 1111–1122. <https://doi.org/10.1016/j.watres.2012.11.033>.
- Sun, B., Lam, D., Yang, D., Grantham, K., Zhang, T., Mutic, S., Zhao, T., 2018. A machine learning approach to the accurate prediction of monitor units for a compact proton machine. *Med. Phys.* 45, 2243–2251.
- Tanimoto, T.T., 1958. Elementary Mathematical Theory of Classification and Prediction. International Business Machines Corp.
- Tomas, O., 2005. A QSAR for the hydroxyl radical reaction rate constant : validation , domain of application , and prediction. *Atmos. Environ.* 39, 2189–2200. <https://doi.org/10.1016/j.atmosenv.2005.01.007>.
- Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciú, M., 2017. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform.* 4, 159–169.
- Wang, Y., Chen, J., Li, X., Wang, B., Cai, X., Huang, L., 2009. Predicting rate constants of hydroxyl radical reactions with organic pollutants: algorithm, validation, applicability domain, and mechanistic interpretation. *Atmos. Environ.* 43, 1131–1135.
- Welbl, J., 2014. Casting random forests as artificial neural networks (and profiting from it). In: German Conference on Pattern Recognition, pp. 765–771.
- Wenzel, J., Matter, H., Schmidt, F., 2019. Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. *J. Chem. Inf. Model.* 59, 1253–1268.
- Xing, J., Wang, H., Luo, K., Wang, S., Bai, Y., Fan, J., 2019. Predictive single-step kinetic model of biomass devolatilization for CFD applications: a comparison study of empirical correlations (EC), artificial neural networks (ANN) and random forest (RF). *Renew. Energy* 136, 104–114.
- Yang, Z.-Y., Dong, J., Yang, Z.-J., Lu, A.-P., Hou, T.-J., Cao, D.-S., 2020. Structural analysis and identification of false positive hits in luciferase-based assays. *J. Chem. Inf. Model.* 60, 2031–2043.
- You, H., Ma, Z., Tang, Y., Wang, Y., Yan, J., Ni, M., Cen, K., Huang, Q., 2017. Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. *Waste Manag.* 68, 186–197.
- Zhong, S., Hu, J., Fan, X., Yu, X., Zhang, H., 2020. A deep neural network combined with molecular fingerprints (DNN-MF) to develop predictive models for hydroxyl radical rate constants of water contaminants. *J. Hazard Mater.* 383, 121141 <https://doi.org/10.1016/j.jhazmat.2019.121141>.
- Zhong, S., Zhang, K., Wang, D., Zhang, H., 2021. Shedding light on “Black Box” machine learning models for predicting the reactivity of HO[rad] radicals toward organic compounds. *Chem. Eng. J.* 405, 126627 <https://doi.org/10.1016/j.cej.2020.126627>.

## 5. CONCLUSIONS AND PERSPECTIVES

In this thesis we used different protocols in order to obtain kinetic parameters, either by means of first principles or based on in silico models. Since the rate constant is a fundamental parameter to interpret the reactivity of chemical reactions, our study permeated different methodologies and provided important mechanisms to understand reactional processes. The main conclusions regarding this work are presented below.:

- i) We employed electronic structure calculations and formulations based on the Transition State Theory to reveal the mechanism, kinetics and toxicity of the picloram reaction with hydroxyl radical.
- ii) We build machine learning models combined with molecular fingerprints to predict the rate constant of organic pollutants in an aqueous environment mediated by OH and  $SO_4^{\cdot-}$  radicals. Finally, we developed a web application to load the models and make predictions of the rate constant in a simple and user-friendly way – it is only necessary to inform the CAS Number or SMILES of the compound.
- iii) We extended the previous protocol for the prediction of the rate constant through the oxidation of atmospheric organic pollutants with OH radical and other kinetic parameters crucial to evaluate the efficiency of the degradation of these compounds.

As a result of the knowledge obtained during the development of this thesis, one of the perspectives in progress are highlighted below:

- extend the models created with other oxidative processes and,
- to evaluate the role of temperature in the prediction of the rate constant.

## 6. REFERENCES

1. Smith, I. W. M. ;*Chem.Soc.Rev.* **2008**, 37, 812.
2. Carvalho-Silva, V. H.; Aquilanti, V.; de Oliveira, H. C. B.; Mundim, K. C. ;*J. Comput. Chem.* **2017**, 38, 178.
3. Laidler, K. J. ;*J. Chem. Educ.* **1984**, 61, 494.
4. Laidler, K. J.; King, M. C. ;*J. Phys. Chem.* **1983**, 87, 2657.
5. Carvalho-Silva, V. H.; Coutinho, N. D.; Aquilanti, V. ;*Molecules* **2020**, 25, 2098.
6. Carvalho-Silva, V. H.; Coutinho, N. D.; Aquilanti, V. ;*Front. Chem.* **2019**, 7, 380.
7. Machado, H. G.; Sanches-Neto, F. O.; Coutinho, N. D.; Mundim, K. C.; Palazzetti, F.; Carvalho-Silva, V. H. ;*Molecules* **2019**, 24.
8. Truhlar, D. G.; Hase, W. L.; Hynes, J. T. ;*J. Phys. Chem.* **1983**, 87, 2664.
9. Laidler, K. ;*Harper Row* **1987**, 1, 233.
10. Bell, R. P. *The Tunnel Effect in Chemistry*; Chapman and Hall: London, 1980.
11. Truhlar, D.; Kohen, A. ;*Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 848.
12. Schmitz, G.; Godtlielsen, I. H.; Christiansen, O. ;*J. Chem. Phys.* **2019**, 150.
13. Sanches-Neto, F. O.; Coutinho, N. D.; Carvalho-Silva, V. H. ;*Phys. Chem. Chem. Phys.* **2017**, 19, 24467.
14. Zhou, Y.; Zhang, D. H. ;*J. Chem. Phys.* **2015**, 194307.
15. Sanches-Neto, F. O.; Coutinho, N. D.; Palazzetti, F.; Carvalho-Silva, V. H. ;*Struct. Chem.* **2020**, 31.
16. Wang, Y. N.; Chen, J.; Li, X.; Zhang, S.; Qiao, X. ;*QSAR Comb. Sci.* **2009**, 28, 1309.
17. Jin, X.; Peldszus, S.; Huck, P. M. ;*Chemosphere* **2015**, 138, 1.
18. Xiao, R.; Ye, T.; Wei, Z.; Luo, S.; Yang, Z.; Spinney, R. ;*Environ. Sci. Technol.* **2015**, 49, 13394.
19. Gupta, S.; Basant, N. ;*RSC Adv.* **2016**, 6, 108448.
20. Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. ;*New J. Phys.* **2013**, 15, 95003.
21. Afzal, M. A. F.; Cheng, C.; Hachmann, J. ;*J. Chem. Phys.* **2018**, 148, 241712.
22. Rupp, M.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. ;*Phys. Rev. Lett.* **2012**, 108, 58301.
23. Houston, P. L.; Nandi, A.; Bowman, J. M. ;*J. Phys. Chem. Lett.* **2019**, 10, 5250.
24. Zhong, S.; Hu, J.; Fan, X.; Yu, X.; Zhang, H. ;*J. Hazard. Mater.* **2020**, 383, 121141.
25. Singh, A. R.; Rohr, B. A.; Gauthier, J. A.; Nørskov, J. K. ;*Catal. Letters* **2019**, 149, 2347.
26. Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. ;*ACS Cent. Sci.* **2018**, 4, 1465.

27. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*; O'Reilly Media, 2019.
28. Sbai, O.; Couprie, C.; Aubry, M. In *European Conference on Computer Vision*; 2020; pp. 597–613.
29. Grambow, C. A.; Pattanaik, L.; Green, W. H. ;*J. Phys. Chem. Lett.* **2020**, *11*, 2992.
30. Komp, E.; Valleau, S. ;*J. Phys. Chem. A* **2020**, *124*, 8607.
31. Zhong, S.; Zhang, K.; Wang, D.; Zhang, H. ;*Chem. Eng. J.* **2021**, *405*, 126627.
32. Jiang, Z.; Hu, J.; Zhang, X.; Zhao, Y.; Fan, X.; Zhong, S.; Zhang, H.; Yu, X. ;*Environ. Res.* **2020**, 109697.
33. Sudhakaran, S.; Amy, G. L. ;*Water Res.* **2013**, *47*, 1111.
34. Lee, Y.; Von Gunten, U. ;*water Res.* **2012**, *46*, 6177.
35. Borhani, T. N. G.; Saniedanesh, M.; Bagheri, M.; Lim, J. S. ;*Water Res.* **2016**, *98*, 344.
36. Cardoso, L. P.; Valim, J. B. ;*J. Phys. Chem. Solids* **2006**, *67*, 987.
37. Yang, Z.; Su, R.; Luo, S.; Spinney, R.; Cai, M.; Xiao, R.; Wei, Z. ;*Sci. Total Environ.* **2017**, *590–591*, 751.
38. Seufert, V.; Ramankutty, N.; Foley, J. A. ;*Nature* **2012**, *485*, 229.
39. Socorro, J.; Durand, A.; Temime-Roussel, B.; Gligorovski, S.; Wortham, H.; Quivet, E. ;*Sci. Rep.* **2016**, *6*, 1.
40. Özcan, A.; Şahin, Y.; Koparal, A. S.; Oturan, M. A. ;*J. Hazard. Mater.* **2008**, *153*, 718.
41. Gao, Y.; Ji, Y.; Li, G.; An, T. ;*Water Res.* **2014**, *49*, 360.
42. An, T.; Gao, Y.; Li, G.; Kamat, P. V.; Peller, J.; Joyce, M. V. ;*Environ. Sci. Technol.* **2014**, *48*, 641.
43. Yang, J.; Wang, Z.; Lv, G.; Liu, W.; Wang, Y.; Sun, X.; Gao, J. ;*Ecotoxicol. Environ. Saf.* **2020**, *197*, 110644.
44. Manonmani, G.; Sandhiya, L.; Senthilkumar, K. ;*J. Phys. Chem. A* **2019**, *123*, 8954.
45. Manonmani, G.; Sandhiya, L.; Senthilkumar, K. ;*Environ. Sci. Pollut. Res.* **2020**.
46. Gao, Y.; Ji, Y.; Li, G.; An, T. ;*Water Res.* **2014**, *49*, 360.
47. Milenković, D. A.; Dimić, D. S.; Avdović, E. H.; Amić, A. D.; Dimitrić Marković, J. M.; Marković, Z. S. ;*Chem. Eng. J.* **2020**, *395*, 124971.
48. Arnaut, L.; Burrows, H. *Chemical kinetics: from molecular structure to chemical reactivity*; Elsevier, Ed.; 1st ed.; Elsevier: The Netherlands, 2006.
49. King, M. C. ;*Ambix* **1984**, *31*, 16.
50. Wilhelmy, L. ;*Pogg. Ann* **1850**, *81*, 422.
51. Wilhelmy, L. ;*Ann. Phys. Chemie* **1850**, *81*, 413.
52. Berthelot, M.; de Saint-Gilles, P. ;*Ann. Chim. Phys.* **1862**, *65*, 385.
53. Guldberg, C. M.; Waage, P. ;*C. Forh. Vidensk. i Christ.* **1864**, *35*, 1864.

54. Guldberg, C. M.; Waage, P. ;*Erdmann's J. fr Pract. Chemie* **1879**, 127, 69.
55. Harcourt, A. V.; Esson, W. ;*Philos Trans R Soc L.* **1866**, 156, 193.
56. Van't Hoff, J. H. *Etudes de dynamique chimique*; Muller, 1884; Vol. 1.
57. Jacobus H. van 't Hoff - Facts - NobelPrize.org  
<https://www.nobelprize.org/prizes/chemistry/1901/hoff/facts/> (accessed Feb 17, 2020).
58. Trautz, M. ;*Chemistry (Easton)*. **1916**, 96, 1.
59. Lewis, W. C. M. ;*J. Chem. Soc. Trans.* **1918**, 113, 471.
60. Fernández-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. ;*Chem. Rev.* **2006**, 106, 4518.
61. Menzinger, M.; Wolfgang, R. ;*Angew. Chemie Int. Ed. English* **1969**, 8, 438.
62. Marcelin, R. ;*J. Chim. Phys.* **1914**, 12, 451.
63. Tolman, R. C. ;*J. Amer. Chem. Soc.* **1920**, 42, 2506.
64. Mills, I.; others *Quantities, units and symbols in physical chemistry/prepared for publication*; Blackwell Science, Ed.; 2nd ed.; Oxford; Boston: Blackwell Science; Boca Raton, Fla.: CRC Press [distributor].: Oxford, 1993.
65. Pacey, P. D. ;*J. Chem. Educ.* **1981**, 58, 612.
66. Eyring, H. ;*J. Chem. Phys.* **1935**, 3, 107.
67. Evans, M. G.; Polanyi, M. ;*Trans. Faraday Soc.* **1935**, 31, 875.
68. Wigner, E. ;*Phys. Rev.* **1932**, 40, 749.
69. Pelzer, H.; Wigner, E. ;*Zeitschrift für Phys. Chemie* **1932**, 15, 445.
70. Peters, B. *Reaction rate theory and rare events*; Elsevier, Ed.; 1st ed.; Elsevier: The Netherlands, 2017.
71. Isaacson, A. D.; Truhlar, D. G.; Rai, S. N.; Steckler, R.; Hancock, G. C.; Garrett, B. C.; Redmon, M. J. ;*Comput. Phys. Commun.* **1987**, 47, 91.
72. CARVALHO, E. F. V. de Estudo teórico da cinética de reações de abstração de hidrogênio do etanol e metanol, Instituto Tecnológico de Aeronáutica, São José dos Campos, 2010.
73. Roberto-Neto, O. Elementos da Dinâmica Química ao nível da Teoria Variacional do Estado de Transição com Correções Interpoladas. *Quim. Nova* **1999**, 22, 737.
74. Carvalho, E. F. V.; Barauna, A. N.; Machado, F. B. C.; Roberto-neto, O. ;*Int. J. Quantum Chem.* **2008**, 108, 2476.
75. Carvalho, V. H.; Aquilanti, V.; de Oliveira, H. C. B.; Mundim, K. C. ;*Rev. Process. Químicos* **2015**, 9, 226.
76. Eckart, C. ;*Phys. Rev.* **1930**, 35, 1303.
77. Bell, R. P. ;*Proc. R. Soc. London. Ser. A, Math. Phys.* **1935**, 148, 241.
78. Bell, R. P. ;*Trans. Faraday Soc.* **1958**, 55, 1.

79. Christov, S. G. ;*Mol. Eng.* **1997**, 7, 109.
80. Aquilanti, V.; Mundim, K. C.; Elango, M.; Kleijn, S.; Kasai, T.; Cavalli, S.; De Fazio, D.; Aguilar, A.; Lucas, J. M.; Coutinho, N. D.; Sanches-Neto, F. O. F. O.; Carvalho-Silva, V. H.; de Oliveira, H. C. B.; Ribeiro, L. A.; Aquilanti, V.; de Oliveira, H. C. B.; Mundim, K. C.; Sanches-Neto, F. O. F. O.; Coutinho, N. D.; Carvalho-Silva, V. H.; Cavalli, S.; Aquilanti, V.; Mundim, K. C.; De Fazio, D. ;*J. Comput. Chem.* **2017**, 498, 24467.
81. Aquilanti, V.; Mundim, K. C.; Elango, M.; Kleijn, S.; Kasai, T. ;*Chem. Phys. Lett.* **2010**, 498, 209.
82. Morel, C. M. ;*Cad. Saude Publica* **2006**, 22, 1522.
83. Silva, V. H. C.; Aquilanti, V.; De Oliveira, H. C. B.; Mundim, K. C. ;*Chem. Phys. Lett.* **2013**, 590, 201.
84. Wigner, E. P. ;*Phys. Rev.* **1948**, 73, 1002.
85. Turing, A. M.; Haugeland, J. *Computing machinery and intelligence*; 1950.
86. Rebala, G.; Ravi, A.; Churiwala, S. In *An Introduction to Machine Learning*; Springer, 2019; pp. 1–17.
87. INTEL Guia de Planejamento Saiba mais sobre Big Data Medidas que Gerentes de TI Podem Tomar para Avançar com o Software Apache Hadoop.
88. ISHIDA, D. M. T. PERSPECTIVAS DO APRENDIZADO DE MÁQUINA NO ENSINO DA ENGENHARIA QUÍMICA, Universidade Federal de São Carlos, 2021.

## PUBLICATIONS, EVENTS AND ACTIVITIES

### ➤ UNRELATED TO THESIS

- **Sanches-Neto, F. O.**, Coutinho, N. D., Palazzetti, F., & Carvalho-Silva, V. H. (2020). Temperature dependence of rate constants for the H (D)+ CH<sub>4</sub> reaction in gas and aqueous phase: deformed Transition-State Theory study including quantum tunneling and diffusion effects. *Structural Chemistry*, 31(2), 609-617.
- Coutinho, N. D., **Sanches-Neto, F. O.**, Carvalho-Silva, V. H., de Oliveira, H. C., Ribeiro, L. A., & Aquilanti, V. (2018). Kinetics of the OH+ HCl→ H<sub>2</sub>O+ Cl reaction: Rate determining roles of stereodynamics and roaming and of quantum tunneling. *Journal of Computational Chemistry*, 39(30), 2508-2516.
- Vieira, C. L., **Neto, F. O. S.**, Carvalho-Silva, V. H., & Signini, R. (2019). Design of apolar chitosan-type adsorbent for removal of Cu (II) and Pb (II): An experimental and DFT viewpoint of the complexation process. *Journal of Environmental Chemical Engineering*, 7(3), 103070.
- Machado, H. G., **Sanches-Neto, F. O.**, Coutinho, N. D., Mundim, K. C., Palazzetti, F., & Carvalho-Silva, V. H. (2019). "Transitivity": a code for computing kinetic and related parameters in chemical transformations and transport phenomena. *Molecules*, 24(19), 3478.
- de Sousa, A. R. V., do Carmo Silva, L., de Curcio, J. S., da Silva, H. D., Eduardo Anunciação, C., Maria Salem Izacc, S., ..., **Sanches-Neto, F. O** & de Paula Silveira Lacerda, E. (2022). "pySewage": a hybrid approach to predict the number of SARS-CoV-2-infected people from wastewater in Brazil. *Environmental Science and Pollution Research*, 1-10.
- Larissa Pinheiro Souza, **Flávio Olimpio Sanches-Neto**, Giberto Mitsuyoshi Yuki Junior, Bruno Ramos, Arlen Mabel Lastre-Acosta, Antonio Carlos Silva Costa Teixeira. Environmental phototransformation of the antidepressant venlafaxine in surface waters: Experimental and theoretical investigation of its persistence in the Billings Reservoir (Sao Paulo, Brazil). Submitted.
- Igor D. Borges, Eduardo C. M. Fariaa, Jean F. M Custódioa, Vitor S. Duarte, Fernanda S. Fernandesc, Christian G. Alonso, Flávio O. Sanches-Neto, Valter H. Carvalho-Silva, Guilherme R. Oliveira and Hamilton B. Napolitano. Insights on chalcone analogues with potential as antioxidant additive for diesel-biodiesel blends. Submitted.
- Bruno Ramos, Wesley F. Vaz, Luan F. Diniz, Flavio Olimpio Sanches Neto, Nayara Dantas Coutinho, Valter H. Carvalho-Silva, Antonio Carlos S. C. Teixeira, Caue Ribeiro, Hamilton B. Napolitano, Paulo de Sousa Carvalho Jr. Kinetics and mechanism of acid hydrolysis degradation of Ametryn and of its tautomeric metabolites in the ionizing states and solid-state structural motifs diversity. Submitted.
- XX Simpósio Brasileiro de Química., Teórica. Flash Talk. Temperature dependence of rate constants for the H(D) + CH<sub>4</sub> reaction in gas and aqueous phase: deformed Transition-State Theory study including quantum tunneling and diffusion. 2019.

- 45 REUNIÃO ANUAL DA SBQ. Pôster. "Synthesis of new 2,5-bis(aryl(chalcogenophen-2-yl)methyl)thiophene derivatives with potencial fluorescent properties." Gabriela Ferreira Matos, LUCILIA KATO, Felipe Lange Coelho, **Flávio Olímpio Sanches Neto**, Jefferson Richard Dias da Silva, Luiz Henrique Keng Queiroz Júnior, Lauro June Queiroz Maia e OLGA SOARES DO REGO BARROS, 2022.
- Substitute Professor at the Federal Institute of Goiás - Campus Anápolis. August 2018 to July 2020.
- Substitute Professor at the Federal University of Goiás. March 2022 to the present.

#### ➤ RELATED TO THESIS

- **Sanches-Neto, F. O.**, Ramos, B., Lastre-Acosta, A. M., Teixeira, A. C. S., & Carvalho-Silva, V. H. (2021). Aqueous picloram degradation by hydroxyl radicals: Unveiling mechanism, kinetics, and ecotoxicity through experimental and theoretical approaches. *Chemosphere*, 278, 130401.
- **Sanches-Neto, F. O.**, Dias-Silva, J. R., Keng Queiroz Junior, L. H., & Carvalho-Silva, V. H. (2021). "pySiRC": Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants. *Environmental Science & Technology*, 55(18), 12437-12448.
- **Sanches-Neto, F. O.**, Dias-Silva, J. R., de Oliveira, V. M., Aquilanti, V., & Carvalho-Silva, V. H. (2022). Evaluating and elucidating the reactivity of OH radicals with atmospheric organic pollutants: Reaction kinetics and mechanisms by machine learning. *Atmospheric Environment*, 275, 119019.
- **Flávio O. Sanches-Neto**, Nayara D. Coutinho, Vincenzo Aquilanti, Wender A. Silva, Valter H. Carvalho-Silva. Degradation by hydroxyl radicals of the potential drugs for the treatment of COVID-19: Unveiling mechanism, kinetics, and ecotoxicity through theoretical approaches. Submitted.
- Encontro Virtual de Materiais e Ciência (e-Mat&Sci). pySiRC: A machine learning combined with molecular fingerprints to predict reaction rate constant of the radical based oxidation processes of aqueous organic contaminants. 2020.
- XXI Simpósio Brasileiro de Química Teórica. Pôster. "DESCRIÇÃO DA CINÉTICA DE DEGRADAÇÃO DE POLUENTES ORGÂNICOS ATRAVÉS DE MÉTODOS DE QUÍMICA QUÂNTICA E DE APRENDIZADO DE MÁQUINA". 2021.
- II Web Encontro Nacional de Engenharia Química, Certificado de Melhor Trabalho. "DESENVOLVIMENTO DE UM MODELO DE APRENDIZAGEM DE MÁQUINA PARA PREDIÇÃO DE CONSTANTES CINÉTICAS DE INTERESSE AMBIENTAL". Gabriel Longo Lopes, Bruno Ramos, **Flávio Olímpio Sanches-Neto** e Antonio Carlos Silva Costa Teixeira. 2022.

- Technical visit at the University of Perugia in Italy in the laboratory of Professor Vincenzo Aquilanti. January 2019.
- Technical visit at the University of São Paulo in the AdOX laboratory. September 2021.

**DECLARAÇÃO DE ORIGINALIDADE DE DISSERTAÇÃO DE MESTRADO OU TESE DE DOUTORADO**

Declaro que a presente dissertação/tese é original, elaborada especialmente para este fim, não tendo sido apresentada para obtenção de qualquer título e que identifico e cito devidamente todas as autoras e todos os autores que contribuíram para o trabalho, bem como as contribuições oriundas de outras publicações de minha autoria.

Declaro estar ciente de que a cópia ou o plágio podem gerar responsabilidade civil, criminal e disciplinar, consistindo em grave violação à ética acadêmica.

Brasília, 10 de outubro de 2022.

Assinatura do/a discente: *Flávio Olímpio Sanches Neto*

Programa: Pós Graduação em Química

Nome completo: Flávio Olímpio Sanches Neto

Título do Trabalho: CINÉTICA DA BIOCUMPLEXIDADE AMBIENTAL:  
EXPERIMENTOS, QUÍMICA QUÂNTICA E APRENDIZADO DE MÁQUINA

Nível: ( ) Mestrado (x) Doutorado

Orientador/a: Valter Henrique Carvalho Silva