On the Existence of Logarithmic Terms in the Drag Coefficient and Nusselt Number of a Single Sphere at High Reynolds Numbers

Yousef M. F. El Hasadi∗1 and Johan T. Padding† 1

1Process and Energy department, Delft University of Technology, Leegwaterstraat 39, 2628 CB Delft The Netherlands

Abstract

In the beginning of the second half of the twentieth century, Proudman and Pearson (JFM, 2(3), 1956, pp. 237-262) suggested that the functional form of the drag coefficient of a single sphere subjected to uniform fluid flow consists of a series of logarithmic and power terms of the Reynolds number. In this paper, we will explore the validity of the above statement for Reynolds numbers up to $2 \times 10^5$, by using a symbolic regression machine learning method. The algorithm is trained by using available experimental data, as well as data from a well-known correlation from the literature. The symbolic regression method finds the following expression for the drag coefficient

∗Email address for correspondence: yme0001@auburn.edu
†Email address for correspondence: J.T.Padding@tudelft.nl
\[ C_D = a + \frac{24}{Re} + f(\log(Re)), \]

where \( Re \) is the Reynolds number, and the constituents of \( f(\log(Re)) \) are integer powers of \( \log(Re) \). Interestingly, the value of \( a \) resembles the value of \( C_D \), at the point where laminar separation point occurs. We did the same analysis for the problem of heat transfer under forced convection around a sphere, and found that the logarithmic terms of \( Re \) and Peclet number \( Pe \) play an essential role in the variation of the Nusselt number \( Nu \). The machine learning algorithm independently found the asymptotic solution of Acrivos and Goddard (JFM, 23(2),pp.273-291).

**Keywords:** sphere, drag coefficient, machine learning, Nusselt number, multi-phase flows, heat transfer, matched asymptotic expansions

### 1 Introduction

Predicting the drag force on an object fixed in a planar flow has been the subject of extensive investigation from the early days of fluid mechanics when it emerged as an independent discipline. The analytical solution for the drag force experienced by a rigid sphere for creeping flow conditions, found by Stokes \[1\] in 1851, is one of the first known analytical expression in the fluid mechanic’s community. Stokes assumed in his solution that inertial effects of the fluid could be neglected throughout the solution domain. However, Oseen \[2\] found an inconsistency in the Stokes solution. Specifically, he found that inertial fluid effects cannot be neglected far away from the sphere. He derived a new form of equations, known as Oseen equations \[2\], that can handle this inconsistency, and he came up with an improved approximation for the drag coefficient, defined as \( C_D = F_D/(\frac{1}{2} \rho v_\infty^2 \frac{\pi}{4} d^2) \), where \( F_D \) is the drag force, \( \rho \) the fluid density, \( v_\infty \) the fluid flow velocity far away from the sphere, and \( d \) the sphere diameter \[3\]. There are additional solutions to the Oseen equations, such those of Goldstein \[4\] and Faxén \[5\].

Proudman and Pearson \[6\] divided the flow field around the sphere into two
stream function expansions. The first one, which they called the Stokes expansion, controls the flow near the surface of the sphere. The second expansion, which they called the Oseen expansion, controls the flow far from the surface of the sphere. Both expansions are based on the Navier-Stokes equations, and the two expansions are matched at a certain distance from the sphere using the method of matched asymptotics. Evaluating stresses from the Stokes expansion they arrived at the following expression for the $C_D$ of a sphere:

$$C_D = \frac{24}{Re} \left(1 + \frac{3}{16} Re + \frac{9}{160} Re^2 \log\left(\frac{Re}{2}\right)\right)$$ (1)

Here $Re = \rho v_{\infty} d/\mu$ is the Reynolds number. They made the following statement (conjecture) about the expansions that govern the flow field [6]: “The non-linearity of the Navier-Stokes equation then shows that both expansions must involve powers of $\log(Re)$, and it seems reasonable to suppose that both expansions are in powers of $Re$, each term of which is multiplied by polynomial in $\log(Re)$”. This statement also, reflects on the functional form of the drag coefficient. However, the authors did not mention the $Re$ range for which the statement is valid. From now on, we will call this conjecture P&P. Graebel [7] supported the P&P statement by mentioning that the $C_D$ functional form that will result from asymptotic expansions of the Navier-Stokes equations will always be a function of $\log(Re)$. A few years later, Chester et al. [8] added an extra term to Eq.(1), which was the last addition that came from the expansion of the Navier-Stokes equations.

The appearance of logarithmic terms (alternatively known as logarithmic switch-back terms [9]) in the asymptotic expansions have intrigued the scientific community, because in some instances they were not forced by the governing equations [10]. Van Dyke [11] dedicated a section in his book describing the proliferation of logarithmic terms in different fluid mechanics problems, and he made the following comment: “one can philosophize that description by fractional powers fails to exhaust the myriad phenomena in the universe, and logarithms are the next simplest
function”. Initially, the logarithms were tied with paradoxes in fluid mechanics, or to the singular perturbation techniques themselves. However, Lagerstrom and Reinelt \[9\] showed that logarithmic terms are part of the solution of the governing equations, and the asymptotic expansion method is just one way to reach to the solution. This view is supported by other investigations using different mathematical methods \[12\,13\].

There are analytical solutions for the Stokes and Oseen regimes for some non-spherical particles such as oblate or prolate spheroids, circular cylinders and few other particle geometries \[14\,17\]. Eq.\,(1) and all other analytical solutions, regardless of the shape of the particles, are valid up to $Re \approx 1.0$. For higher $Re$, analytical solutions for the Navier-Stokes equations cease to exist due to its nonlinearity. For the prediction of $C_D$ at higher $Re$ we usually resort to correlations that have been fitted to either experimental or numerical simulations. For the drag on a sphere, there are plenty of correlations that take different mathematical forms \[18\,25\], as shown in the extensive list published in the recent review by Goossens \[26\]. The majority of correlations take the following functional form:

$$C_D = \frac{24}{Re} (C_1 + C_2Re^a) + \frac{C_3}{1 + \frac{Re}{Re_c}}$$

(2)

The second term of Eq.\,(2) is coming from boundary layer theory \[27\], which accounts for the inertial effects of the fluid. The value of the exponent $a$ ranges from 0.5 to 0.68. These type of correlations are suitable for $Re$ as high as $2 \times 10^5$, right before the so-called drag-crisis.

Concerning the heat transfer rate from a particle fixed in a fluid, most investigations available in the literature are related to the case of forced convection. In this type of flow, the velocity profile is decoupled from that of the temperature. For further simplification, there is also no variation in the transport properties of the fluid with temperature. These simplifications pave the way of obtaining sev-
eral analytical solutions for a single sphere \[28\] for limited cases of low \( Re \) and \( Pe = v_\infty d/\alpha \), where \( \alpha \) is the thermal diffusivity of the fluid. Acrivos and Taylor \[28\] used asymptotic expansions and the velocity profile of the Stokes solution to find the following relation for the Nusselt number \( Nu = h d/k \), where \( h \) is the (convective and surface mean) heat transfer coefficient and \( k \) is the thermal conductivity of the fluid (linked to the thermal diffusivity through \( k = \alpha \rho c_p \), with \( c_p \) the specific heat capacity of the fluid), for the case of \( Pe \to 0 \) and \( Re \to 0 \):

\[
Nu = 2 + \frac{1}{2}Pe + \frac{1}{4}Pe^2 \log(Pe) + 0.034Pe^2 + \frac{1}{16}Pe^2 \log(Pe)
\] (3)

In practice, this solution is limited to \( Re \lesssim 0.03 \). Rimmer \[29\] added an extra term to Eq.(3) from asymptotic expansions, and as far as we know this is the last term that evolved from the matched asymptotic expansions in the low \( Pe \) and \( Re \to 0 \) regime. Conversely, for \( Pe \to \infty \) and \( Re \to 0 \), Acrivos and Goddard \[30\] used the matched asymptotic expansions to arrive at the following relation for \( Nu \):

\[
Nu = 0.922 + 1.249Pe^\frac{1}{3}
\] (4)

As for the case of drag, for higher \( Re \) we need to rely on semi-empirical relations to express the variation of \( Nu \) with the flow field parameters. Whitaker \[31\] provided a correlation, which is still considered one of the most accurate available in literature \[32\]:

\[
Nu = 2 + (C_4Re^{a_1} + C_5Re^{a_2})Pr^{a_3}
\] (5)

Where \( Pr = c_p\mu/k \) is the Prandtl number (note that \( Pe = RePr \)). The values of \( a_1, a_2, \) and \( a_3 \) are \( \frac{1}{2}, \frac{2}{3}, \) and \( 0.4 \), respectively. The Whitaker correlation is valid for \( 1 \leq Re \leq 10^5 \) and a wide range of \( Pr \). The second, and third terms represent inertial fluid effects, and their functional form is inspired by boundary layer theory. Although the first term comes from the analytical solution for pure conduction from a sphere, all exponents in Eq.(5) are obtained from empirical fitting.
In summary, almost all correlations for drag and heat transfer found in literature are expressed as power law expansions, similar to Eqs. (2), (4) and (5). Correlations with logarithmic terms, such as Eqs. (1) and (3), are extremely rare and seem to have been largely overlooked.

In this paper we will use symbolic regression, which is a modern tool for unbiased determination of correlations, to re-investigate known data on drag and heat transfer. We will show that symbolic regression actually rediscovers the logarithmic terms, suggesting that logarithmic expansions may represent the physics better than power law expansions. As a side result, we will show that there is an intriguing connection between the found logarithmic terms and the point of first boundary layer separation.

2 Methodology

In this paper, we will use the symbolic regression machine learning method proposed by Koza [33]. Symbolic regression is a powerful tool for searching the mathematical space for an approximate functional relation between a certain number of input and output variables, and it is based on genetic programming proposed by Holland [34]. The framework of genetic programming is probabilistic, and is not based on mathematical principles, such as correctness, consistency, justifiability, certainty, orderliness, and decisiveness as outlined by Koza [33], but solely on the principles of Darwinian evolution [35]. The idea of the genetic programming is simple, and it is based on transforming an initial population (in our case a population of mathematical functions) to a new population that survived a particular fitness constraint. The main operators that are used to create the new population are similar to those found in nature, namely that of reproduction and crossover [33].

The algorithm first generates a random pool of functions, that undergo genetic
operations such as crossover, which corresponds to the combination of two functions to give a new offspring function. Another operation is a mutation in which a certain part of the mathematical function is changed randomly. Two indexes measure the fitness of the newly obtained functions. The first index is minimizing the mean square difference between the training and predicted dependent values. The second index is to check the mathematical complexity of functions, and select the simplest ones, to prevent over-fitting. We used the Eureqa software \cite{36} as symbolic regression platform. A rigorous description of the symbolic regression algorithm in use in the current investigation is given in \cite{37}.

3 Results

In the first subsection, we will explore the dependence of the drag coefficient $C_D$ on $Re$ for a fixed sphere. We will devote the second subsection to explore the dependence of the Nusselt number $Nu$ of a sphere on $Re$ and $Pe$ (or $Pr$) for the case of forced convection with constant transport properties.

3.1 Drag coefficient $C_D$

We will start by exploring the $C_D$ dependency on $Re$ for the case of a sphere. We will create three data sets for the regression process. The first one will be generated from the correlation of Brown and Lawler \cite{38} which has the functional shape of Eq.\,(2). This data set contains about 8500 points in the range $0.1 < Re < 1.9 \times 10^5$, which is enough to capture the smallest details in the $C_D$ variation. The second data set that we will use is the exact experimental data that Brown and Lawler \cite{38} used themselves to derive their correlation. It contains about 450 points in the range $0.1 < Re < 1.975 \times 10^5$. The final data set is based on the Schiller and Naumann \cite{39} correlation, and contains 5020 points in the range $0.1 < Re < 700$.

We will start by examining the first data set, and we will let the symbolic
regression algorithm guess about the functional form of the $C_D$ dependence on $Re$. We can do this by specifying the most general initial functional form:

$$C_D = f(Re)$$  \hspace{1cm} (6)

The algorithm derived several regression equations, but here we will show two, one because it accurately fits the results, and the other because it is simple. The equations are the following:

$$C_D = a_1 + \frac{a_2}{Re} + a_3\sqrt{Re} + \frac{a_4}{\sqrt{Re}} + \frac{a_5}{(a_6 + Re)} + a_7Re$$  \hspace{1cm} (7)

$$C_D = a_1 + \frac{a_2}{Re} + \frac{a_3}{\sqrt{Re}}$$  \hspace{1cm} (8)

The coefficients of Eq.(7), and Eq.(8) are listed in Table 1. The structure of Eq.(8) contains the Stokes $\frac{1}{Re}$ term, and the first-order term from boundary layer theory $\frac{1}{\sqrt{Re}}$. The first known dependency of $C_D$ on $\frac{1}{\sqrt{Re}}$ came from the Blasius solution [40] of the boundary layer equations proposed by Prandtl [41] for the case of a flat plate. The $C_D$ for blunt bodies, like a sphere, has a similar dependency on $Re$ [42, 43]. A similar form as Eq.(8) was obtained previously by fitting experimental data [44, 45], and also by using concepts of boundary layer theory [43]. Refs [44,45] used non-linear fitting tools to obtain their correlations, which require a priori knowledge of the functional structure. A comparison between the the coefficients of Eq.(8), and those of Refs [43–45] is given in Table (2). The coefficients of Eq.(8) have similar values to those of [44]. Compared to those of [45] there is only significant difference in the value of $a_3$. There is also a significant difference between the coefficients of Eq.(8) and those of Abraham [43]. This may be due to the pure theoretical nature of the equation proposed by Abraham.

It is important to note that both the Stokes term and the boundary layer term have been found without using any sophisticated mathematical approach. On the contrary, they have been found by a probabilistic genetic algorithm. The emergence of the boundary layer term in Eqs. (7) and (8) without human intervention
can be added to the experimental and numerical results that support boundary layer theory even though there is no general mathematical proof of its existence, as mentioned by Batchelor [46].

We will now try to explore the existence of logarithmic switchback terms for the drag on a sphere for the higher $Re$ regime. We will use for this the first data-set (i.e. data from the Brown and Lawler [38] correlation). We will start by imposing the following initial functional form:

$$C_D = f\left(\frac{24}{Re}, \log(Re), Re \log(Re), \log^2(Re)\right)$$ (9)

We choose this form of the initial function because we want to ensure that logarithmic switchback terms similar to Eq.(1) will be part of the initial soup of functions that the symbolic algorithm will further evolve. The symbolic regression algorithm converged to the following equation:

$$C_D = a_1 + \frac{a_2}{Re} + a_3 \log(Re) + a_4 \log^2(Re) + a_5 \log^4(Re)$$ (10)

The values of the coefficients of Eq.(10) are listed in Table 3. Eq.(10) depends on powers of $\log(Re)$ and also contains the Stokes law term. The form of Eq.(10) is partially fulfilling the P&P conjecture [6] for $Re$ as high as $2 \times 10^5$. Overall, Proudman and Pearson [6] made a profound statement more than 64 years ago using only mathematical intuition, and they may have been right when they suspected that logarithmic switchback terms are part of the solution. It may be difficult for the current form of the genetic algorithm to spot the entire logarithmic switchback series, because reducing the complexity of the equations is part of its optimization process. Therefore, terms that do not play a significant role in the variation of the dependent variable ($C_D$) will die out during the evolution process. The failure of detection of $Re^n \log^n(Re)$ terms, where $n$ is an integer, after a significant number of mathematical formula evaluations exceeding $10^{11}$, suggests that their signal is weak (a metaphor for their insignificant role in the dependence of $C_D$ on $Re$). If
we read more carefully the conjecture, we find that Proudman and Pearson [6] used the following wording: “It seems reasonable to suppose that both expansions are in powers of Re”. They used the word ‘reasonable to suppose’, expressing doubt, while for the log(Re) terms they used the word ‘must’ which reflects that the authors were sure about their appearance in the two expansions. Adding to that, Proudman [8] was frustrated about the poor convergence of his equation, mainly because it is only valid for extremely low values of Re. He suggested that the expansion in powers of Re may be a poor idea [8, 47].

In order to further validate the ecosystem of the equations that we obtained, we will compare their predictions with various sources in the literature, as shown in Figure 1. The first insight from Figure 1 is that Eq. (1) is valid only at low Re, and this was one of the main reasons we believe that the scientific community did not further explore the use of logarithmic terms, even as fitting functions. Eq. (7) and Eq. (10) follow closely the correlation of Brown and Lawler [38], and also the experimental data used to obtain the correlation of [38]. The average relative errors between the predictions of Eq. (7), and Eq. (10) with respect to the experimental results of [38] are 3.87% and 3.39%, respectively. We see that Eq. (8) follows closely the results of [44, 45], while it deviates from the predictions of Abraham [43] especially for values of Re above 10^3. This is expected because the equation provided by Abraham [43] is valid for Re up to 10^3. Also, Eq. (8) and those of References [43–45] cannot capture the local minimum for Re between 10^3 and 10^4 that the experimental results of [38] show.

Comparing Eq. (7) and Eq. (10), we find that their complexity index is 34 and 19 respectively. The complexity index shows that the logarithmic series representation of CD is mathematically simpler compared to the power series representation, making Eq. (10) more favourite to represent the physical phenomena of the CD variation according to the Occam razor statements [48]. One of these statements
is: “Given two models with the same generalization error, the simpler one should be preferred because simplicity is desirable in itself.”

Now we will use the second (experimental) data set, to explore the feasibility of getting predictive equations for $C_D$ from a limited amount of noisy experimental data. We will start by letting the algorithm guess the $C_D$ dependence:

$$C_D = f(Re)$$  \hspace{1cm} (11)

The symbolic regression algorithm found the following equation:

$$C_D = a_1 + \frac{a_2}{Re} + \frac{a_3}{\sqrt{Re}}$$  \hspace{1cm} (12)

The coefficients of Eq.(12) are listed in Table 1. Using the second data set we next explore if the data show any logarithmic dependence by imposing the following initial set of functions:

$$C_D = f(\frac{24}{Re}, \log(Re), Re \log(Re), \log^2(Re))$$  \hspace{1cm} (13)

We got the following equation for $C_D$:

$$C_D = a_1 + \frac{a_2}{Re} + \frac{a_3 \log^2(Re)}{Re} + a_4 \log(Re) + a_5 \log^2(Re)$$  \hspace{1cm} (14)

The values of the coefficients are listed in Table 3. Eq.(12) is of a similar form as Eq.(8), but the coefficients are not identical, because the second data set contains far less data, and also contains some noise. The derivation of Eq.(12) from pure experimental data, without imposing knowledge of any physics, except the definition of $Re$, shows that the symbolic regression algorithm discovered the Stokes formula, and the term attributed to boundary layer theory without any external help. The algorithm needed less than an hour to discover what took human intellect hundreds of years to achieve. However, the human factor is still required since we have to select the equations that we think represent physical reality from the population of equations that the algorithm suggests. Eq.(14) shows that we can
get the logarithmic dependence from a pure experimental data set, and it partially fulfils the P&P conjecture. Eq. (14) and Eq. (11) are quite similar. We believe that Eq. (14) could not capture the $\log^4(Re)$ term because this term influences $C_D$ in the high $Re$ regime where there are significant fluctuations in the experimental data set. Probably if there were a higher volume of data, especially at higher $Re$, the $\log^4(Re)$ term could also be captured from pure experimental results. A comparison of the performance of Eq. (12) and Eq. (14) against existing data in the literature is shown in Figure 2. The average relative error for Eq. (12) and Eq. (14) is 13.7% and 12.0%, respectively, against the experimental results of [38]. Eq. (14) shows a local minimum in the range of the $Re$ close to that of the experimental results of [38], while Eq. (12) fails to show any local minimum.

We will use the third and final data set from the Schiller and Naumann [39] correlation which contains information about the variation of $C_D$ for $Re$ ranging from 0.1 to 700. We will use the following general initial functional form:

$$C_D = f(Re)$$

(15)

The symbolic regression algorithm found the following equation for $C_D$:

$$C_D = a_1 + \frac{a_2}{Re} + a_3 \log(Re) + a_4 \log^2(Re)$$

(16)

The coefficients of Eq. (16) are listed in Table 4. The genetic algorithm came up with the logarithmic dependence of $C_D$ on $Re$ without any external help, and it discovered the P&P conjecture partially. The value of $a_1 = 3.1406$, differs from the value of $\pi$ by only about 0.03%. It will be very interesting in the future to investigate the value of $a_1$ from fitting very accurate numerical or experimental data. Eq. (16) follows the Brown and Lawler correlation [38] up to $Re$ of $10^3$, as shown in Figure 1. This behaviour is expected because the higher power logarithmic terms are missing from Eq. (16), since the training data is limited to $Re$ up to 700.
Up to this point we have discussed the drag without referring to the flow around the sphere. The flow around a sphere is a rich mosaic of phenomena, and usually drag correlations, fail to predict them. Among these phenomena is the emergence of a laminar separation point, which is well known to occur for sufficiently blunt objects including a sphere. The point of laminar separation is identified by the formation of a closed recirculating ring eddy at the rear of the sphere. The first emergence of separation is difficult to detect either experimentally or theoretically. For this reason, there is some discrepancy in the reported critical $Re_s$, and corresponding drag $C_{D_s}$, in the literature. The first experimental observations by Nisi and Porter [49] suggested that $Re_s = 10$. This was confirmed by numerical simulations of Rimon and Cheng [50]. On the other hand, Proudman and Pearson [6], and Van Dyke [11], by using the Stokes second expansion, estimated that $Re_s = 16$, close to the numerical results of Bourot [51] and Jenson [52] of 15.2 and 17, respectively, and the experiments of Payard and Countanceau [53] indicating $Re_s = 17$. Other simulation results [54, 55] show that $Re_s$ is equal to approximately 20, and the experiments of Taneda [56] predict that $Re_s = 24$.

If we inspect $a_1$ of Eq.(10) in Table 3 we see that its value is 3.286, which is quite similar to the value of $C_{D_s}$ at the initial laminar separation reported by [53], which is 3.306. If the constant $a_1$ is the drag coefficient at initial laminar separation, then the following transcendental equation must have a positive root at the corresponding Reynolds number $Re_s$:

$$\frac{a_2}{Re} + a_3 \log(Re) + a_4 \log^2(Re) + a_5 \log^4(Re) = 0 \quad (17)$$

By solving Eq.(17) we found that $Re_{rt} = 14.06$ is its only root. That makes $Re_{rt}$ the only $Re$ value that zeroes off all terms beyond the constant $a_1$. This $Re_{rt}$ is close to values of $Re_s$ reported in literature. For example, the relative error with respect to the results of Bourot [51] and Chang and Maxey [54] is 8% and 30%, respectively. We conjecture that $Re_{rt}$ is representing $Re_s$, even though we do
not have any proof for this. We believe we are witnessing an instance where the machine learning algorithm found a mathematical description of a physical phenomenon, which needs human abilities to be interpreted in terms of physical laws. Otherwise, it will be a good approximation, that can describe some of the physics involved in the process of flow separation. As far as the authors are aware, there is only one analytical prediction for the point of first flow separation, from slow motion viscous theory [3, 57]. However, that result was disputed by the authors of [3, 57], as we will show later. In practice, we depend on numerical simulations to find the point of zero local shear stress, as described by boundary layer theory [27]. However, Batchelor [46] raised serious doubts about estimating the onset of separation by this method.

Beyond this point, we will assume that (the smallest, real) root $Re_s$ is equal to $Re_s$. Using the same procedure to calculate $Re_s$, from Eq.(14) by solving the following transcendental equation:

$$\frac{a_2}{Re} + \frac{a_3 \log^2(Re)}{Re} + a_4 \log(Re) + a_5 \log^2(Re) = 0$$

(18)

we found the two following roots: $Re_s = 15.76$, and $9.52 \times 10^7$. The large root value of $9.52 \times 10^7$, is a non-physical result, which we believe is caused by the missing higher power $\log(Re)$ term from Eq.(14). However, $Re_s = 15.76$ compares very well with the results of Bourot [51] and Chang and Maxey [55], with a relative difference of 3.68% and 21.2%, respectively. If we do the same analysis for Eq.(16), we will find that $Re_s = 15.19$, and $3.518 \times 10^6$. For the smallest root, the relative difference with the results of Bourot [51] and Chang and Maxey [55] is 0.13%, and 24.0%, respectively.

We will next calculate $Re_{rt}$ from the more popular power-law expressions Eq.(7) and Eq.(8) in the same way. For Eq.(7) we find the following roots $Re_{rt1} = -2461 - 767i$, $Re_{rt2} = -2461 + 767i$, and $Re_{rt3} = 3 \times 10^5$. The first two roots are
non-physical, while the third root is the result of the divergence of Eq. (17) beyond the value of $Re = 2 \times 10^5$, which is the limit of the training data. We certainly believe that $Re_{rt3}$ does not convey any physical significance. As for Eq. (18), it does not have any roots, neither in the real nor in the complex domain.

Returning to the logarithmic ecosystem of equations, in their seminal works, Proudman and Pearson [6] and Van Dyke [11] calculated the $Re_s$ value to be 16 analytically from the first and second terms in the Stokes expansion. Proudman and Pearson [6] made the following comment: "This Reynolds number is far too large to make estimates based on only two terms of the Stokes expansion at all reliable. In fact, it cannot seriously be claimed that slow-motion theory gives even a qualitative expansion of the phenomena." However, Van Dyke [11] and Ranger [58] tried to confirm the result of Proudman and Pearson [6], by using extra terms in the Stokes expansion that contain the logarithmic terms from the results of Proudman and Pearson [6] and those of Chester et al. [8]. They failed because the Stokes expansion equation that includes the logarithmic terms has only complex roots. Van Dyke [11] commented on this issue saying that "the logarithm needs reinterpretation." In our work we now see that the values of $Re_s$ from Eq. (10), Eq. (14), and Eq. (16) are converging with different degree of accuracy toward a value of approximately 16.

In summary, in this section we showed that the functional form of $C_D$ could be represented by either powers or logarithmic functions of $Re$. However, the logarithmic representation conveys the physics in a different way than the power representation, and illuminates new physical phenomena, which are beyond the reach of current analytical, or empirical $C_D$ formulas. When appealing to mathematical aesthetics, our results suggest that the drag coefficient of a sphere might be well described by the form $C_D = \pi + 24/Re + f(\log Re)$, with $C_D = \pi$ at the first point of separation, occurring at a Reynolds number $Re_s$ given by $24/Re_s + f(\log Re_s) = 0$. 
Van Dyke [11] described the appearance of logarithms in the asymptotic expansions as obscure, but it appears that these obscure entities can speak the language of fluid dynamics much better than powers. A similar situation exists in the field of turbulence, especially regarding channel flow, where there is an open debate in the scientific community whether power or logarithmic expansions bests describe the velocity at the wall in certain flow regimes [59]. Note that the logarithmic dependence of the drag coefficient $C_D$ also exists for geometries different than a sphere such as spherocylinders, and prolate spheroids, as shown by our previous work [37].

### 3.2 Nusselt number $Nu$

In this section, we will explore the possibly logarithmic dependence of the Nusselt number $Nu$ on the Peclet number $Pe$, and Reynolds number $Re$. For this purpose we will create a data set of 26,796 points from the Whitaker [31] correlation Eq.(5) for $Pr$ in the ranging from 0.74, to 7.0, and $Re$ in the range of $10^{-1}$ to $10^4$. We will start with the simplest assumption by allowing the symbolic regression algorithm to guess about the dependency of $Nu$ on $Re$, $Pr$ and/or $Pe$, through the following initial function:

$$Nu = f(Re, Pr, Pe)$$

(19)

The resulting $Nu$ correlation is the following:

$$Nu = a_1 + a_2 \sqrt{Pe} + a_3 \sqrt{Re} \sqrt{a_4 + a_5 \sqrt{Pe} + a_6 Pe + a_7 Re}$$

(20)

The coefficients are listed in Table 5. Most of the equations that the algorithm produces show that $Nu$ is a function of $Re$ and $Pe$, and excludes the dependence on $Pr$. This is different from the source of our data (the Whitaker correlation Eq.(5)), which explicitly depends on $Pr$ and $Re$. Even when we used a substantial amount of data, the algorithm failed to predict the exact structure of the Whitaker correlation [31]. The recent investigation of Udrescu and Tegmark [60] showed,
consistent with our results, that Eureqa failed to predict the exact functional structure of many functions included in the Feynman lectures [61]. They attributed this failure due to the complexity of those functions, and the number of variables that they contain.

Examining the properties of Eq. (20), we find that as \( Re \to 0 \), Eq. (20) reduces to \( a_1 + a_2 \sqrt{Pe} \), which bears similarities with Eq. (4) for the \( Pe \) dependency, because for both cases the power of \( Pe \) is less than one, and both equations show that even at very low \( Re \) the convection affects the heat transfer rate. This type of dependency did not exist in the Whitaker correlation Eq. (5), where for \( Re \to 0 \) (outside the range of validity of the Whitaker correlation) \( Nu \) converges to a value of 2.0, corresponding to pure conduction from a single sphere.

We will now examine the full dependence of \( Nu \) on the logarithms of \( Pe \), \( Re \), and \( Pr \). This structure of dependency is based on our previous knowledge of the physics of the problem of forced convection over a sphere. We know that for \( Re \to 0 \) and \( Pe < 1 \), \( Nu \) depends on \( \log(Pe) \) \[28\] (Eq. 3), so there may exist an intermediate \( Pe \) regime where logarithms will play a role as well, until we reach a high \( Pe \) regime where Eq. (1) is dominant. For the high \( Re \) regime we already showed that the drag coefficient \( C_D \) is a function of logarithms of \( Re \), so because of the tight relation between flow and heat transfer [62] we expect that logarithms of \( Re \) will play a role in the convective heat transfer process as well. The initial function has the following form:

\[
Nu = f(\log(Pe), Pe \log(Pe), \log^2(Pe), \log(Re), Re \log(Re), \\
\log^2(Re), \log(Pr), Pr \log(Pr), \log^2(Pr))
\] (21)

As initial guess we gave equal weight to all the independent variables functional forms, to avoid any bias, toward any of the independent variables. The symbolic
regression algorithm found the following two correlations:

\[ \text{Nu} = a_1 + a_2 \log^2(Re) \log(Pe) Pe^{a_3} + a_4 Pe^{a_5} \quad (22) \]

\[ \text{Nu} = a_1 + a_2 \log^2(Re) + a_3 Pe^{a_4} + a_5 \log^2(Re) \log(Pe) Pe^{a_6} + a_7 \log(Pe) \quad (23) \]

The second equation is more complex than the first. The coefficients of both Eq. (22), and Eq. (23) are listed in Table 5. Both equations possess very interesting features. We will start with Eq. (23), where the term \( a_1 + a_3 Pe^{a_4} \) resembles closely the approximation of Eq. (4). The relative difference of the \( a_1 \), \( a_3 \) coefficients and those of Eq. (4) is 15\%, and 8\%, respectively. The relative error is remarkably small, if we take into account that the source of the data set is coming from an empirical correlation that has an average predictive error of 30\%.

We believe that the combination of the logarithmic dependence of \( Pe \) and \( Re \) plays an essential role in the emergence of an asymptotic solution. It seems there are very few possible ways to fit the data of [31] using logarithms of \( Pe \) and \( Re \) and one of those few is using terms similar to Eq. (4). This also explains why Eq. (20) failed to predict accurately the original source of the training data, Eq. (4). Our findings show the essential role played by previous physical knowledge of the problem in specific regimes, to help the machine learning algorithm to reach a physically meaningful result.

The genetic algorithm predicted the asymptotic solution for the high \( Pe \) (Eq. 4) case, rather than for low \( Pe \) (Eq. 3), probably because our training data is more biased toward the high \( Pe \) regime. Since the lowest \( Re \) and \( Pr \) used are 0.1, and 0.7 respectively, the lowest \( Pe \) we used is 0.07, which lies at the boundary of the high \( Pe \) regime. We could not use lower \( Pe \) because the Whitaker correlation [31] is based on \( Re \) ranging between 3.5 and \( 7.6 \times 10^4 \), and \( Pr \) ranging between 0.7
and 380. Note that we did use the Whitaker correlation [31] also for lower Re, 0.1 < Re < 3.5, to generate our training data. We test its validity against the experimental data of Will et al. [63] for the lowest Prandtl number that we used, Pr = 0.7, and for Re as low as 0.1, and we found that the Whitaker correlation [31] follows closely the results of [63], as shown in Figure 3. An indication that the hydrodynamics in the highly inertial regime may be governed by logarithmic terms of Re, is the the appearance of log^2(Re) both in Eq.(22) and Eq.(23), similar to the case of CD (see Eqs.(10),(14) and (16)). Also, the log^2(Re) terms for both Nu and CD share the same sign, and their pre-factors are of the same order of magnitude.

We compare the performance of our predictor equations for different Pr, and Re numbers, in Figure 3. We select four cases, two of them lie within the training data set (Pr = 0.7 and 7.0) that we supplied to the algorithm. The other two test cases (Pr = 50 and 300) lie outside the training data set to test the extrapolation capabilities of our predictor equations. For Pr = 0.7, Eqs (20), (22) and (23) perfectly follow the Whitaker [31] correlation and the experimental results of Will et al. [63]. At high Re they also follow the numerical results of Feng and Michaelides [64]. As expected, our ecosystem of equations do not follow the asymptotic solution of Acrivos and Goddard [30] since their solution is only valid in the low Re and high Pe regime. For the case of Pr = 7.0, our ecosystem of equations predicts the evolution of Nu with great accuracy. For the cases of Pr = 50, and 300, Eqs.(22) and (23) predict with great accuracy the results of the Whitaker [31] correlation, except in a very narrow region at low Re. The conditions in this low Re, but at the same time high Pr regime are applicable to the asymptotic solution of Acrivos and Goddard [30]. This why the whole ecosystem of our equations deviate from the results of the Whitaker [31] correlation, and follow by different degrees of accuracy the asymptotic solution of Acrivos and Goddard [30], Eq.(4). All of our equations are functions of Pe and Re. However, for low Re the Nu cor-
relations switch to a dependency on $Pe$ only, which is consistent with the physics of Eq. (3) and (4).

The above shows that symbolic regression can find an asymptotic solution by using previous physical knowledge, rather than depending completely on the training data set. Feeding machine learning algorithms previous physical knowledge for the problem that they try to optimize, increases substantially the probability of better extrapolation predictions. For further discussion on how to implement previous knowledge into symbolic regression, the readers is referred to our recent publication [37].

4 Conclusions

In this investigation, we explored the possibility of a logarithmic dependence of the drag coefficient $C_D$, and Nusselt number $Nu$ on the Reynolds number $Re$ and Peclet number $Pe$, inspired by the asymptotic solutions for the creeping flow conditions. We used a symbolic regression machine learning algorithm, and our training data are based on experiments, and data from well-known empirical correlations available in the literature. We can make the following conclusions:

- The drag coefficient $C_D$ can be expressed as a function of powers in $\log(Re)$ partially fulfilling the Proudman and Pearson [6] conjecture P&P.

- If an expansion in terms of $\log(Re)$ is made for the drag coefficient $C_D$, the value of the $Re$ at which all the $Re$ dependent terms go to zero is closely resembling $Re$ at the first emergence of value of the laminar separation, as predicted analytically by Proudman and Pearson [6].

- The logarithmic dependence of $C_D$ on $Re$ is found independently, without any prior knowledge, by the symbolic regression algorithm.
• The Nusselt number of a single sphere depends on logarithms of $Re$, $Pe$, as well as powers of $Pe$.

• If logarithmic functions of $Re$ and $Pe$ are used as initial functions for the symbolic regression algorithm, the algorithm produces with high accuracy the asymptotic solution derived by Acrivos and Goddard [30] from the matched asymptotic method, at low $Re$, and high $Pe$ regime. Interestingly, the training data that we used does not follow the asymptotic solution of Acrivos and Goddard [30].

The bigger picture of our results shows that if one day we manage to solve in a closed form the Navier-Stokes equations, combined with the heat equation around a sphere, the most probable outcome is to find logarithms rather than powers in the solution. We admit that our method cannot give answers as rigid as mathematical proofs, but only probable answers.

**Acknowledgements**

The first author thanks Dimitra Damianidou for the enlightening discussions about the subject. The authors thank Lorenzo Botto for the discussion about the matched asymptotic methods. Finally, the authors thank the European Research Council for its financial support under its consolidator grant scheme, contract no. 615096 (NonSphereFlow).

**References**

[1] Stokes, G. G. *On the effect of the internal friction of fluids on the motion of pendulums*, volume 9. Pitt Press Cambridge, Cambridge, UK, (1851).

[2] Oseen, C. W. *Arkiv Mat., Astron. och Fysik* 6, 1 (1910).
[3] Oseen, C. W. *Ueber den gueltigkeitsbereich der stokesschen widerstandsformel*. Friedländer, (1913).

[4] Goldstein, S. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 123(791), 225–235 (1929).

[5] Faxen, H. *Arkiv for Matematik Astronomi och Fysik*. 17, 1–28 (1923).

[6] Proudman, I. and Pearson, J. *Journal of Fluid Mechanics*. 2(3), 237–262 (1957).

[7] Graebel, W. *Advanced fluid mechanics*. Academic Press, Burlington, MA, USA, (2007).

[8] Chester, W., Breach, D., and Proudman, I. *Journal of Fluid Mechanics*. 37(4), 751–760 (1969).

[9] Lagerstrom, P. and Reinelt, D. *SIAM Journal on Applied Mathematics*. 44(3), 451–462 (1984).

[10] Popović, N. In *Journal of Physics: Conference Series*, volume 22, 164. IOP Publishing, (2005).

[11] Van Dyke, M. *NASA STI/Recon Technical Report A*. 75 (1975).

[12] Holzer, M., Kaper, T. J., et al. *Advances in Differential Equations*. 19(3/4), 245–282 (2014).

[13] Popović, N. and Szmolyan, P. *Nonlinear Analysis: Theory, Methods & Applications*. 59(4), 531–565 (2004).

[14] Happel, J. and Brenner, H. *Low Reynolds number hydrodynamics: with special applications to particulate media*, volume 1. Springer Science & Business Media, (2012).
[15] Cox, R. *Journal of Fluid Mechanics* **23**(4), 625–643 (1965).

[16] Breach, D. *Journal of Fluid Mechanics* **10**(2), 306–314 (1961).

[17] Aoi, T. *Journal of the Physical Society of Japan* **10**(2), 119–129 (1955).

[18] Rouse, H. Technical report, Dover Publications Inc, New York, USA, (1961).

[19] Engelund, F. and Hansen, E. *A monograph on sediment transport in alluvial streams*. TEKNISKFORLAG Skelbrekgade 4 Copenhagen V, Denmark., (1967).

[20] Clift, R. *Proc. Chemeca’70* **1**, 14 (1970).

[21] Morsi, S. and Alexander, A. *Journal of Fluid mechanics* **55**(2), 193–208 (1972).

[22] Graf, W. H. *Hydraulics of sediment transport*. Water Resources Publication, Highlands Ranch, CO USA, (1984).

[23] Flemmer, R. L. and Banks, C. *Powder Technology* **48**(3), 217–221 (1986).

[24] Khan, A. and Richardson, J. *Chemical Engineering Communications* **62**(1-6), 135–150 (1987).

[25] Swamee, P. K. and Ojha, C. S. P. *Journal of Hydraulic Engineering* **117**(5), 660–667 (1991).

[26] Goossens, W. R. *Powder Technology* **352**, 350–359 (2019).

[27] Schlichting, H. and Gersten, K. *Boundary-layer theory*. Springer, Berlin Heidelberg, Germany, (2016).

[28] Acrivos, A. and Taylor, T. D. *The Physics of Fluids* **5**(4), 387–394 (1962).

[29] Rimmer, P. L. *Journal of Fluid Mechanics* **32**(1), 1–7 (1968).
[30] Acrivos, A. and Goddard, J. *Journal of Fluid Mechanics* **23**(2), 273–291 (1965).

[31] Whitaker, S. *AIChE Journal* **18**(2), 361–371 (1972).

[32] Sparrow, E. M., Abraham, J. P., and Tong, J. C. *International Journal of Heat and Mass Transfer* **47**(24), 5285–5296 (2004).

[33] Koza, J. R. *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press, Cambridge, MA, USA, (1992).

[34] Holland, J. H. et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, Cambridge, MA, USA, (1992).

[35] Darwin, C. *On the origin of species*. Johon Murray, London, UK, (1859).

[36] Schmidt, M. and Lipson, H. *science* **324**(5923), 81–85 (2009).

[37] El Hasadi, Y. M. and Padding, J. T. *AIP Advances* **9**(11), 115218 (2019).

[38] Brown, P. P. and Lawler, D. F. *Journal of environmental engineering* **129**(3), 222–231 (2003).

[39] Schiller, L. and Naumman, A. *Z. Vereines Deutscher Inge.* **77**, 318–321 (1933).

[40] Blasius, H. *Grenzschichten in Flüssigkeiten mit kleiner Reibung*. University of Gottingen, Germany, (1908).

[41] Prandtl, L. In *Verhandlg. III. Intern. Math. Kongr. Heidelberg*, 484–491. Math Congress, (1904).

[42] Leal, L. G. *Advanced transport phenomena: fluid mechanics and convective transport processes*, volume 7. Cambridge University Press, Cambridge, UK, (2007).
[43] Abraham, F. F. *The Physics of Fluids* **13**(8), 2194–2195 (1970).

[44] Brauer, H. and Mewes, D. *Chemie Ingenieur Technik* **44**(13), 865–868 (1972).

[45] Hölzer, A. and Sommerfeld, M. *Powder Technology* **184**(3), 361–365 (2008).

[46] Batchelor, G. *An introduction to fluid dynamics*. Cambridge university press, Cambridge, UK, (2000).

[47] Hunter, C., Tajdari, M., and Boyer, S. *SIAM Journal on Applied Mathematics* **50**(1), 48–63 (1990).

[48] Domingos, P. *Data mining and knowledge discovery* **3**(4), 409–425 (1999).

[49] Nisi, H. and Porter, A. W. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **46**(275), 754–768 (1923).

[50] Rimon, Y. and Cheng, S. *The Physics of Fluids* **12**(5), 949–959 (1969).

[51] Bourot, J. *A* **269**, 1017–1020 (1969).

[52] Jenson, V. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **249**(1258), 346–366 (1959).

[53] Payard, M. and Coutanceau, M. *CR Academie des Sciences, Paris B* **278**, 369–372 (1974).

[54] Dennis, S. and Walker, J. *Journal of Fluid Mechanics* **48**(4), 771–789 (1971).

[55] Chang, E. J. and Maxey, M. R. *Journal of Fluid Mechanics* **277**, 347–379 (1994).

[56] Taneda, S. *Journal of the Physical Society of Japan* **11**(10), 1104–1108 (1956).

[57] Van Dyke, M. *Journal of Fluid Mechanics* **44**(2), 365–372 (1970).

[58] Ranger, K. *SIAM Journal on Applied Mathematics* **23**(3), 325–333 (1972).
[59] Schultz, M. P. and Flack, K. A. *Physics of Fluids* **25**(2), 025104 (2013).

[60] Udrescu, S.-M. and Tegmark, M. *arXiv preprint arXiv:1905.11481* (2019).

[61] Feynman, R. P., Leighton, R. B., and Sands, M. *American Journal of Physics* **33**(9), 750–752 (1965).

[62] Duan, Z., He, B., and Duan, Y. *Scientific reports* **5**, 12304 (2015).

[63] Will, J. B., Kruyt, N. P., and Venner, C. H. *International journal of Heat and Mass transfer* **109**, 1059–1067 (2017).

[64] Feng, Z.-G. and Michaelides, E. E. *International Journal of Heat and Mass Transfer* **43**(2), 219–229 (2000).
| Coefficients | Eq.(7) | Eq.(8) | Eq.(12) |
|--------------|--------|--------|---------|
| $a_1$        | 0.251  | 0.412  | 0.505   |
| $a_2$        | 23.620 | 23.311 | 23.224  |
| $a_3$        | 0.001  | 4.119  | 2.762   |
| $a_4$        | 3.255  | -      | -       |
| $a_5$        | 49.291 | -      | -       |
| $a_6$        | 97.537 | -      | -       |
| $a_7$        | $-2.709 \times 10^{-6}$ | - | - |

Table 1: Coefficients for Eq.(7), Eq.(8), and Eq.(12)

| Coefficients | Ref [44] | Ref [45] | Ref [43] |
|--------------|----------|----------|----------|
| $a_1$        | 2.9%     | -1.94%   | 29.01%   |
| $a_2$        | -2.95%   | -2.95%   | -2.87%   |
| $a_3$        | 2.88%    | 27.16%   | -28.40%  |

Table 2: Relative difference in the values of coefficients of Eq.(8) to that of Brauer and Mewes [44], Holzer and Sommerfeld [45], and Abraham [43].

| Coefficients | Eq.(10) | Eq.(14) |
|--------------|---------|---------|
| $a_1$        | 3.286   | 3.272   |
| $a_2$        | 24.205  | 23.26   |
| $a_3$        | -0.818  | 0.112   |
| $a_4$        | 0.064   | -0.652  |
| $a_5$        | -0.000107 | 0.035   |

Table 3: Coefficients for Eq.(10) and Eq.(14)
Table 4: Coefficients for Eq.(16)

| Coefficients | Eq.(16) |
|--------------|---------|
| $a_1$        | 3.140   |
| $a_2$        | 24.270  |
| $a_3$        | -0.716  |
| $a_4$        | 0.047   |

Table 5: Coefficients for Eq.(20), Eq.(22), and Eq.(23)

| Coefficients | Eq.(20) | Eq.(22) | Eq.(23) |
|--------------|---------|---------|---------|
| $a_1$        | 2.0     | 1.582   | 1.063   |
| $a_2$        | 0.343   | 0.003   | 0.0067  |
| $a_3$        | 0.0454  | 0.326   | 1.351   |
| $a_4$        | 9.341   | 1.0     | 0.299   |
| $a_5$        | 1.0     | 0.322   | 0.0028  |
| $a_6$        | $-7.0 \times 10^{-5}$ | -     | 0.332   |
| $a_7$        | -0.00131 | -     | -0.128  |

Table 5: Coefficients for Eq.(20), Eq.(22), and Eq.(23)
Figure 1: Comparison between the drag coefficient $C_D$ predicted by Eq.(7), Eq.(8), Eq.(10), Eq.(16) and different sources from the literature. Dashed lines indicate literature correlations. Symbols indicate experimental values.
Figure 2: Comparison between drag coefficient $C_D$ predicted by Eq. (12) Eq. (14), and different sources from the literature. Dashed lines indicate literature correlations. Symbols indicate experimental values.
Figure 3: Comparison between the results of different predictor equations for the Nusselt number \( Nu \) with those from literature for four different Prandtl numbers \( Pr \).