aerospace

MDPI

Article

Symbolic Regression for Air Transport Delay Analysis: A Viable Alternative to Classical Approaches?

Massimiliano Zanin

Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Parc Bit, 07120 Palma, Spain; mzanin@ifisc.uib-csic.es

Abstract

Delays are among air transport's main operational challenges, with significant economic, societal and environmental consequences, and many methodological alternatives have been used in their study. Here we explore the use of symbolic regression, a data-driven technique that searches a space of analytic expressions to identify compact and interpretable models explaining a given set of data. We specifically use symbolic regression to characterise delays at the busiest European airports, how they evolve in time and depend on their own past, up to how they propagate across airports. This is done with the aim of evaluating the feasibility of using this approach, and the added value when compared to standard statistical and causal models. Results of this proof of concept point to a nuanced picture: while symbolic regression demonstrates clear potential for uncovering interpretable functional relationships in delay dynamics, its applicability is hindered by the significant computational cost and its stochastic nature.

Keywords: air transport; delays; symbolic regression; Granger causality

1. Introduction

Delays are among the most relevant and persistent challenges in air transport, affecting passengers, airlines, and airport operators alike. Delays generate substantial direct costs through additional fuel burn, crew overtime, and ground handling charges, as well as indirect costs arising from missed connections, reduced passenger confidence, and reputational damage to carriers [1,2]. At a systemic level, the problem is further compounded by the tendency of delays to propagate; reactionary (or secondary) delays amplify initial perturbations far beyond the initial event. Given this complexity, the scientific study of air transport delays has drawn on a broad methodological repertoire. These include, on the one hand, descriptive statistics and queuing models, aimed at characterising distributions and at identifying their primary drivers [3–6]. On the other hand, the community has progressively been adopting more sophisticated data-driven approaches, including time series analysis [7,8], network-theoretic methods [9–13], causality tests [14–19], and, more recently, machine learning and Deep Learning techniques [20–24]. Each of these tools has contributed to a richer understanding of delay dynamics. At the same time, the outcomes can be difficult to interpret and constrained by strong assumptions about the form of the underlying patterns to be detected.

Among the many alternatives available in the data analysis literature, one of the most conceptually interesting is symbolic regression (SR), i.e., the supervised learning task aimed at finding the optimal analytic expression to describe a set of data. The origins of SR date back to the 1970s and the development of the BACON system by Pat Langley,

Check for updates

Academic Editor: Lorenzo Castelli

Received: 6 May 2026
Revised: 1 June 2026
Accepted: 5 June 2026
Published: 8 June 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

i.e., a program designed to rediscover empirical laws from tabular data by iteratively applying algebraic heuristics. Examples included the recovery of Kepler's third law, Coulomb's law, or Ohm's law from raw measurements [25]. This was then followed by other solutions, including DALTON, GLAUBER, or STAHL [26]. These were nevertheless limited by the high computational cost, a consequence of the large size of the search space. A quantitative jump only happened in the next decade, with the introduction of genetic programming, and the formal coinage of the term "symbolic regression" [27]. This further evolved, until the demonstration of the feasibility of autonomously recovering Hamiltonians, Lagrangians, and conservation laws from motion-tracking data of physical systems, without any theoretical prior knowledge [28].

Note that this review of the history of SR has been kept intentionally brief; at the same time, interested readers can refer to multiple works and books on the topic [29–31]. Still, across the aforedescribed evolution, the central goal has remained constant: to discover compact and interpretable symbolic expressions directly from data, without imposing a predetermined model structure. The focus on interpretability is what makes SR stand out from other data analysis alternatives, and especially from black-box machine learning and Deep Learning models.

In spite of the potential advantages that SR can bring to air transport, the number of works at the intersection between both fields has hitherto been scant. Most relevant examples can be found in the adjacent fields of aerodynamics and turbulence modelling, as for instance in Refs. [32–35], and of aircraft engine health monitoring [36], and mechanical fatigue in general [37]. Also, SR has been used in the study of delays, but only for train, as, e.g., in Ref. [38], where it was used to discover algebraic expressions for delay distributions from Czech railway data.

In this contribution we explore whether SR can be used to model delays in air transport and their propagation. We specifically consider two scenarios. Firstly, the past history of delays at one airport is used to forecast their future evolution; this can thus be used as a proxy for assessing the presence of memory in the dynamics, or whether, on the contrary, delays are purely stochastic values. Secondly, we evaluate whether past delays at one airport can help in predicting the future evolution of delays at a second one. This is equivalent to a predictive causality analysis [39], and is thus similar to what is measured by the celebrated Granger Causality test [40,41], but without enforcing a specific analytic form in the forecasting model. In both cases, these analyses are not novel, but rather reflect questions that have been tackled in the literature—see for, instance, respectively Refs. [4,42,43] and [18,44,45].The novelty here resides in the use of SR to answer them. Specifically, we here present a proof of concept regarding the application of SR in this context. We thus aim at assessing whether it yields a clear added value compared to more traditional approaches, and whether such advantages outpace the complexity of the analysis.

The remainder of this contribution is organised as follows. Section 2 describes the data and methods used in the analysis, including the real-world delay data (Section 2.1) and the symbolic regression framework (Section 2.2). Sections 3 and 4 then present the results obtained for the two scenarios here considered, respectively the self-prediction of delay time series at individual airports, and cross-airport prediction as a proxy for delay propagation. Finally, Section 5 discusses the implications of these results, evaluates the added value of symbolic regression with respect to more conventional approaches, and outlines possible directions for future work.

2. Materials and Methods

2.1. Delay Data

Delay data were extracted from EUROCONTROL's Aviation Data Repository for Research, a public repository of historical flights made available for research purposes and freely accessible at https://www.eurocontrol.int/dashboard/aviation-data-research (accessed on 1 October 2025). It includes information about all commercial flights operating in and over Europe, as well as their flight plans and radar trajectories. A key limitation of this repository is that it provides data for only four months each year, i.e., March, June, September, and December; the selection of these four months is thus not a decision of the author, but instead a constraint at the source.

The arrival delay of each flight has been calculated as the difference between the actual and the scheduled landing times, the latter encoded in the last-filed flight plan. Delays of individual operations have then been aggregated to obtain time series encoding the average hourly landing delay at a given airport. In order to avoid the abnormal dynamics caused by COVID-19, time series have been limited to the period March 2015–December 2019. Finally, in order to ensure that enough operations are available for each airport, while also limiting the computational cost of the analysis, the top-20 airports, in terms of number of operations, have here been considered.

2.2. Symbolic Regression

Symbolic regression (SR) refers to a supervised learning task aimed at finding the optimal analytic expression to describe a set of data. Specifically, we start with two sets of data, respectively denoted by x and y, for which we suppose a relation $y = f(x)$ exists; the objective is then to find the optimal approximation of f. Note that this could be in principle tackled with two opposite approximations. Firstly, one may try to fit the parameters of some over-parametrised general model; while a solution can be found, it would come at the cost of a high risk of overfitting, and would further have low interpretability. Secondly, classical machine learning models try to solve the same problem, but building upon an underlying hypothesis about the structure of such relationship—e.g., linear Support Vector Machines assume that two groups of instances are linearly separable [46]. In contrast to these approaches, SR searches the space of simple analytic expressions, jointly minimising prediction error and model complexity.

To illustrate the basic idea behind SR, suppose $x = \{2, 3\}$ and $y = 6$. The approach involves testing a set of simple equations on these numbers, in order to select the one minimising the error. For instance, these may be $x_1 + x_2 = 5$ (with an error of 1), $x_1 x_2 = 6$ (the best answer, with an error of 0), and so on. These equations are created by combining basic operations on the input x, using some criteria for exploring the whole space. Additionally, simple expressions are preferred over more complicated ones; e.g., even if $2x_1 + x_2 - x_2/x_2$ is technically valid, it would be discarded in favour of the simpler $x_1 x_2$.

Due to the infinite dimensionality of the space of analytic expressions, one essential ingredient of any SR implementation is the algorithm used to explore it. In the last decades, many alternatives have been proposed, based on concepts that include Deep Learning [47], Bayesian statistics [48], or Monte Carlo trees [49]. In this work we use the implementation included in the PySR Python package (version 1.5.9) [50], an open-source symbolic regression framework that uses evolutionary algorithms to discover interpretable equations. PySR's internal search algorithm is a multi-population evolutionary algorithm, which consists of an evolve–simplify–optimise loop: it firstly explores the expression space through genetic operations, including mutation and crossover; redundant or unnecessary operations are removed; and finally, constants are optimised using the classical BFGS optimisation algorithm [51]. Additionally, the algorithm operates on a Pareto frontier,

maintaining a set of non-dominated solutions across different complexity levels. In this work, input data corresponding to landing delays at airports are combined through the set of basic operators listed in Table 1. Unless otherwise specified, the complexity of the solution (i.e., the maximum number of operators applied to the data) is set to seven, and the training is performed for 10^3 iterations over 200 individuals.

Table 1. Operators considered in this work, classified by type. Relu and erf respectively correspond to the Rectified Linear Unit and the Error functions.

Type	Operators
Unary operators	cos, sin, exp, $\log_2$, relu, erf
Binary operators	$+, -, \times, /$

3. Analysis of Individual Airports

We start the analysis of delays by considering each airport separately; in other words, the objective is to model the evolution of delays at a given airport from their recent history. Given $d_i(t)$, representing the average landing delay at airport i and at time t, SR is used to infer the function $d_i(t) = f[d_i(t-1), d_i(t-2), \ldots, d_i(t-6)] + \xi$, i.e., the best estimation given the last six values of d; ξ represents the residuals of the estimation.

Table 2 reports the best function obtained for the top-20 European airports, with the corresponding mean squared error (MSE) in the third column—see also the end of the manuscript for a list of airport codes. In order to estimate whether the result is significant, the last two columns report the MSE obtained when randomly shuffling the time series of each airport, hence destroying any temporal structure in it, and the Z-Score of the MSE against that of the random shuffling.

Table 2. Analysis of the dynamics of the delays at individual airports. Each row reports the best function approximating the average delay observed at time t, given the delay in previous time steps. The subsequent columns report the mean squared error of the approximation (MSE); the same error, calculated on randomly shuffled time series (rMSE); and the corresponding Z-Score. Airport codes are listed in the Abbreviations Section.

Airport	$d(t)$=	MSE	rMSE	Z-Score
EGLL	$2866.4 \sin(d(t-1)/2953.3)$	923,917	1,194,536	−24.91
LFPG	$659.51 \cos(-\cos(d(t-1)/835.82))$	366,611	374,345	−11.55
EHAM	$relu(d(t-1))$	328,457	581,255	−15.81
EDDF	$relu(d(t-1))$	236,559	333,209	−9.62
LEMD	$0.24\, d(t-1) + 244.7$	425,536	444,047	−2.39
LEBL	$relu(d(t-1))$	269,897	409,199	−11.94
EDDM	$0.30\, d(t-1) + 140.4$	185,971	201,343	−4.64
EGKK	$relu(d(t-1)\, erf(d(t-2)))$	1,114,729	1,661,682	−2.69
LIRF	$0.19\, (d(t-4) + d(t-1))$	311,055	319,162	−4.53
LFPO	$0.63\, d(t-1)$	185,009	245,960	−13.70
EIDW	$0.36\, d(t-1)$	283,445	314,252	−16.00
LSZH	$0.31\, d(t-1) + 202.91$	203,443	222,616	−14.40
EKCH	$d(t-1)/6.66$	158,183	160,043	−5.14
LEPA	$relu(d(t-1))$	251,191	334,020	−11.72
LPPT	$relu(0.65\, d(t-1))$	344,055	434,945	−13.62
ENGM	$0.27\, d(t-1)$	301,199	316,442	−6.39
EGCC	$relu(0.44\, d(t-1)) + 225.33$	372,992	402,244	−15.70
EGSS	$0.45\, d(t-1) + 250.37$	290,065	342,221	−18.31
LOWW	$0.19\, d(t-1) + 314.71$	410,424	423,151	−4.15
ESSA	$0.20\, d(t-1) + 135.30$	210,976	217,236	−5.60

It can be appreciated that, in most cases, delays are a linear function of those observed one hour before. A few airports display small non-linearities around that linear behaviour, including trigonometric and relu functions. This latter case is especially interesting, as $relu(x)$ for $x < 0$ is equal to zero; in other words, negative delays do not propagate in time. This is consistent with what is expected in real operations: the early arrival of an aircraft does not affect the next flight, as the subsequent departure time remains unchanged. For the sake of clarity, the functional relationships $d(t) = f[d(t-1)]$ inferred by the SR algorithm are also graphically depicted in Figure 1. Deviations of the fitted curves (blue lines) from the diagonal (grey lines) reflect the degree and directionality of delay attenuation or amplification over consecutive time steps. In all cases, except for EGKK (London Gatwick), delays at time t only depend on what was observed at time $t-1$; if longer correlations were present, their magnitude is too small to be detected by this approach. Additionally, all Z-Scores are negative, confirming that the derived functions are actually describing temporal relationships in the delay time series.

Figure 1 further reports the detected relationships when only data for summer months (i.e., June and September) are used—see the orange lines. Except for a few exceptions (e.g., Paris Charles de Gaulle, LFPG, and London Gatwick, EGKK), the relationships between past and future delays are similar; thus, they seem to reflect limited seasonality patterns.

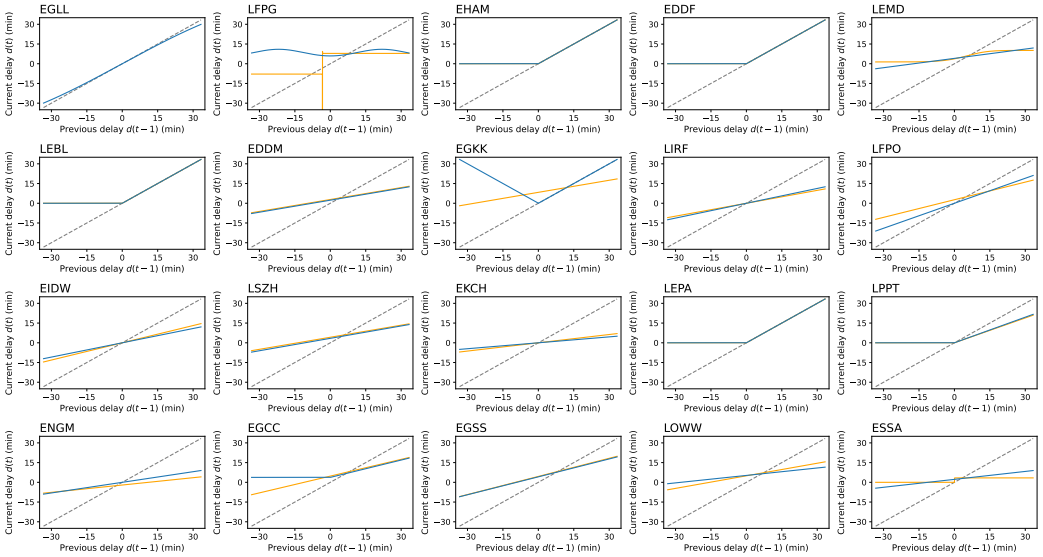


Figure 1. SR models of delay evolution for the top-20 European airports. Each panel displays the functional relationship $d(t) = f[d(t-1)]$ inferred by the SR algorithm; see also Table 2. Blue and orange lines respectively correspond to all data, and data for the months of June and September alone. The dashed grey lines indicate the identity $d(t) = d(t-1)$. Airport codes are listed in the Abbreviations Section.

Due to the stochastic nature of the SR algorithm used in this work, and the fact that the global optimum is not necessarily achieved by the underlying genetic algorithm optimiser, we next evaluate the stability of these results. Specifically, Figure 2 reports a characterisation of the SR solution obtained for EGLL, London Heathrow. The SR optimisation was here repeated 200 times, and the corresponding results have been evaluated. The best-fit equation always takes the form $d(t) = C_2 \cdot \sin[d(t-1)/C_1]$ in all realisations. Left and centre panels of Figure 2 display the empirical distributions of respectively C_1 and C_2 across these realisations. Both distributions are sharply concentrated around their modal values ($C_1 \approx 2953$ and $C_2 \approx 2866$), indicating that the inferred equation is robust. Next, the right panel of Figure 2 reports the mean (right axis, green) and standard deviation (left axis, blue) of the MSE across runs as a function of the number of evolutionary iterations. As is to be

expected, both decrease with increasing iterations, as the solution becomes better, but also more stable across realisations.

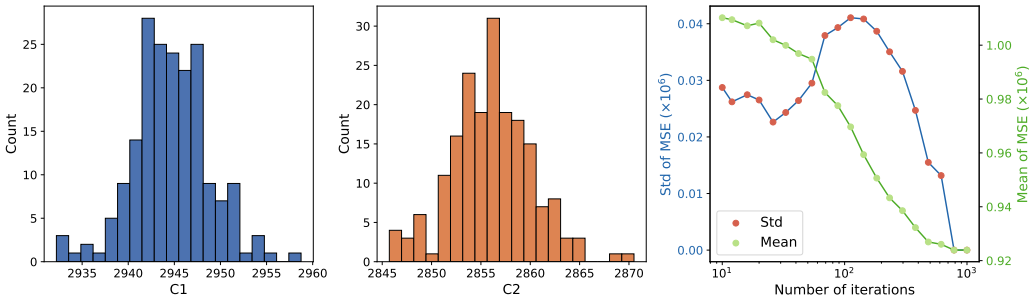


Figure 2. Stability of the SR model for EGLL (London Heathrow). From left to right, panels report the distribution of C_1, of C_2, and of the mean and standard deviation of the MSE as a function of the number of iterations. Results correspond to 200 independent realisations of the SR process.

While Figure 2 illustrates that results yielded by the method are stable, a natural question that emerges is whether the underlying dynamics are also stable, or, on the contrary, if airports are not stationary, with delays evolving according to rules that change with time. To tackle this question, Figure 3 reports the temporal evolution of the MSE across the five-year observation period for three major European airports, namely EGLL (London Heathrow), LFPG (Paris Charles de Gaulle), and EHAM (Amsterdam Schipol). The time series of hourly average delays at each airport has been divided into twenty equally sized time windows, and an independent SR optimisation has been executed on each of them. Solid lines in Figure 3 then represent the mean MSE, while shaded bands indicate the corresponding standard deviation.

An interesting observation emerges from this analysis. Both LFPG and EHAM present a stable evolution of the MSE, with relatively minor peaks; the same is nevertheless not true for EGLL, with a notable spike in the MSE around June 2017. In other words, the SR approach fails at providing a good forecasting model for that month, suggesting that the underlying dynamics have become less predictable. This may partly be explained by the British Airways global IT outage of 27 May 2017, which caused all flights of this airline from Heathrow and Gatwick to be cancelled for approximately two days, with rippling consequences in the following days. This was also followed by a second IT failure in August 2017, which may have affected delays in September of that year. Yet, it is worth noting that the abnormal spike in the MSE had started already in March of that year, suggesting that more causes may be behind it.

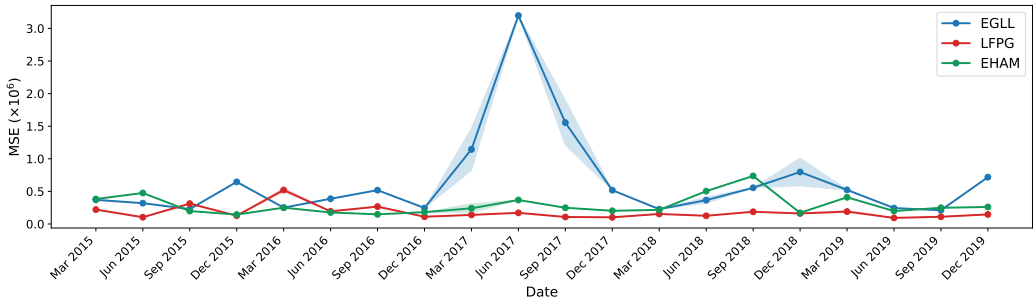


Figure 3. Dynamics of airports through time. The plot depicts the evolution of the mean MSE (solid lines), and of the corresponding standard deviation (shaded bands), for EGLL (London Heathrow), LFPG (Paris Charles de Gaulle), and EHAM (Amsterdam Schipol). Results correspond to 20 independent realisations.

As a last point in this section, we evaluate how the SR approach performs, in terms of the error in the prediction of future delays, compared to other standard machine learning approaches. This was performed by splitting the data set corresponding to EGLL into two halves. The first half was used to train an SR model on the one hand, and a Gradient Boosting (GB) model [52,53] and a Random Forest (RF) regressor [54,55] on the other. The latter two are state-of-the-art models commonly used in many real-world applications, which yield high accuracies while requiring a fraction of the computational cost of more complex Deep Learning models, yet they share a similar problem of low explainability. These models were then applied to the second half of the data, and the squared error achieved by them was calculated for each available data point. Results are presented in Figure 4 as two scatter plots. Notably, the three models perform in a similar way; the average squared errors are respectively 5.07% higher and 13.27% lower for GB and RF, as compared to the SR; i.e., the two machine learning models do not provide any major improvement. Additionally, Figure 4 indicates that the highest variability corresponds to data points with a low error; conversely, some values are difficult to predict for all models, suggesting that they correspond to abnormal and unpredictable disturbances.

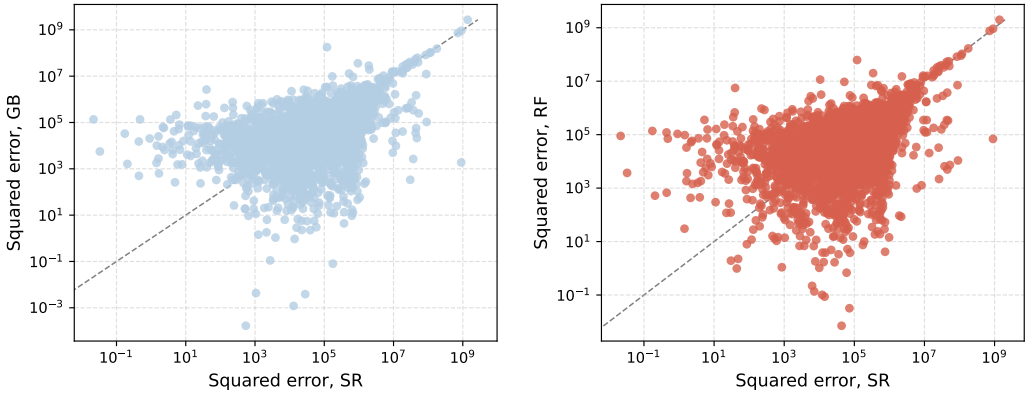


Figure 4. Comparison of the predicting power of SR against machine learning models. The two scatter plots report the squared error in the prediction of individual values, for a Gradient Boosting model (GB, **left** panel) and a Random Forest regressor (RF, **right** panel), as a function of the same error obtained by the SR approach. Dashed lines represent the main diagonal.

4. Analysis of Delay Propagation Between Airports

The natural extension of the previous analysis can be obtained by considering pairs of airports; the aim then becomes describing how the delays at one airport can be explained both by their history, as well as by the history of delays at the second aerodrome. Mathematically, we want to find the function $d_i(t) = f[d_i(t-1), \ldots, d_j(t-1), \ldots] + \xi$ that minimises the residuals ξ, considering both past delays at i and j. In a way similar to what was previously presented in Section 3, we evaluate the statistical significance of the results by comparing the MSE with the one obtained when the delays at airport j are randomly shuffled. This destroys any temporal relationship between the two airport dynamics, and hence allows us to estimate the real contribution of airport j to the future of i.

Table 3 reports a list of the top-ten relations in term of Z-Score, or, in other words, the ten pairs of airports (i, j) for which the past of j has the strongest influence on the future of i. On the other hand, the Z-Scores for all pairs of airports are reported in the left panel of Figure 5. Several interesting differences can be observed when comparing these results with the ones in Table 2. Firstly, most of the relationships are here non-linear, the only exceptions being the pairs EDDM $\rightarrow$ LFPG and EDDF $\rightarrow$ LFPG—see Figure 6 for a graphical representation. In most cases, such non-linearity appears at $d_j(t-1) = 0$; i.e., what is transmitted is mainly the sign of the average delay. Secondly, the functional form

of the relation seems to be conserved (or at least, is highly similar) across pairs sharing the same destination airport. To illustrate, both aforementioned relations targeting LFPG have a linear form with similar coefficients. It is also noteworthy that, for many airport pairs, the found relationship includes dependencies on d_j but not on d_i; in other words, the future delay evolution at one airport depends on the past of a different one, more than on its own past, which may suggest the influence of the global status of the system. Finally, while the SR model is fed with the last six time points for both airports, i.e., from $d(t-6)$ to $d(t-1)$, in most cases only the information for $t-1$ is actually used.

Table 3. Analysis of the relationships between pairs of airports. d_i and d_j respectively represent the delays of the airport to be explained (or the destination), and of the explaining airport (the source). Columns additionally report MSEs and Z-Scores, as in Table 2. Airport codes are listed in the Abbreviations Section.

Source	Destination	Equation	MSE	rMSE	Z-Score
EDDF	LSZH	$\exp(6.02\,erf(d_j(t-1)))$	199,602	234,105	−31.09
EDDF	LOWW	$240.28/\cos(erf(d_j(t-1)))$	409,969	425,604	−14.75
EDDM	LFPG	$0.18\,d_j(t-1)+400.55$	375,623	382,648	−10.75
EDDF	LFPG	$0.19\,d_j(t-1)+385.24$	372,717	381,726	−8.02
LSZH	EGLL	$d_i(t-1)\,erf(d_j(t-1)^2)$	904,237	1,227,787	−5.91
LFPO	EGLL	$d_i(t-1)\,erf(d_j(t-1)^2)$	917,219	1,294,352	−5.63
LSZH	LEMD	$d_i(t-1)\,relu(erf(d_j(t-2)))$	407,774	461,389	−5.47
EGLL	LIRF	$relu(d_i(t-1))\,\sin(0.0005\,d_j(t-2))$	300,818	310,930	−4.67
EDDM	LOWW	$d_j(t-1)/3.46+329.10$	409,724	424,019	−3.88
LSZH	LIRF	$d_i(t-1)\,\sin(relu(erf(d_j(t-1))))$	302,063	353,367	−3.86

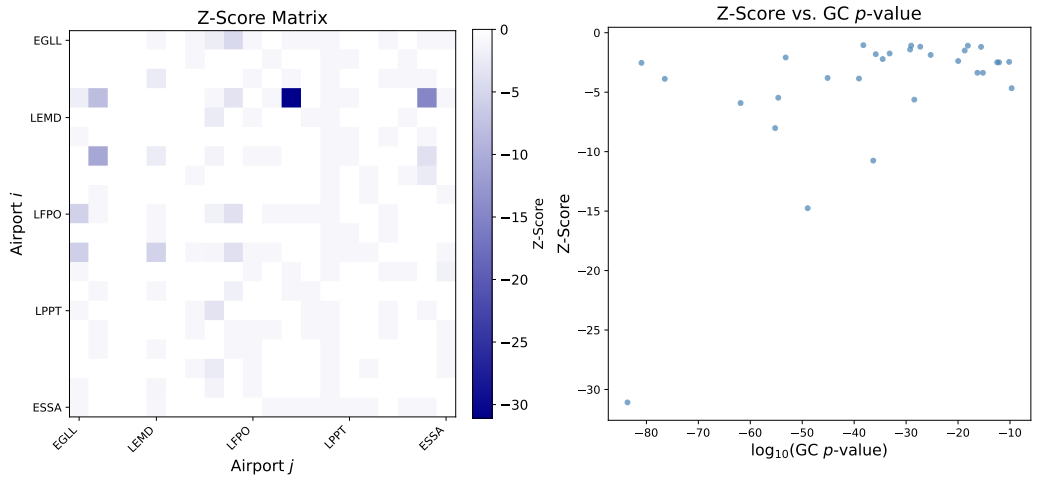


Figure 5. Relationships between pairs of airports. (**Left**) Z-Score obtained for each pair of airports. In both axes, airports follow the same order as in Table 2. (**Right**) Z-Score obtained through the SR approach as a function of the p-value yielded by the Granger Causality test. Each point corresponds to a pair of airports for which the Z-Score is negative.

It is worth noting that this analysis is conceptually similar to a Granger Causality test, according to which a causality relation is present whenever the past of the causing element improves the prediction of the future of the caused element [40,41]. The main difference resides in the way the prediction is performed: a linear auto-regressive model (ARMA) in the case of the Granger test, as opposed to a non-linear function in the case at hand. In order to compare both approaches, the right panel of Figure 5 reports a scatter plot of the Z-Score for pairs of airports with a significant relationship (i.e., Z-Score < 0), as a function of the p-value yielded by a Granger Causality test. A weak agreement can be appreciated, with more negative Z-Scores usually corresponding to smaller p-values. This is further

confirmed by Pearson's ($\rho = 0.5198$, p-value of 3.24×10^{-3}) and Spearman's correlations ($\rho = 0.3967$, p-value of 3.00×10^{-2}).

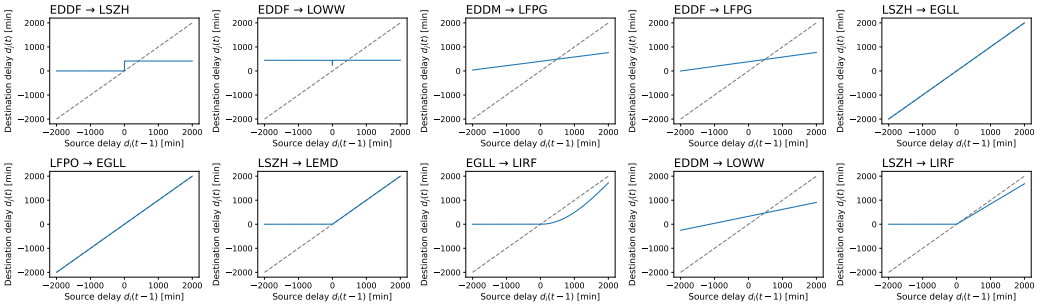


Figure 6. SR models of delay propagation between airport pairs. Each panel reports the functional relationship $d_i(t) = f[d_j(t-1)]$ inferred by the SR algorithm for the top-10 airport pairs; see Table 3. Blue lines represent the fitted relationships, while dashed grey lines indicate the identity $d_i(t) = d_j(t-1)$. The title of each panel indicates the direction of influence as source → destination. Airport codes are listed in the Abbreviations Section.

Delving deeper into these results, the connection between EDDF and LSZH is the strongest one, both according to SR and Granger Causality. In order to verify its resilience, we performed a local search around the only constant in the function (i.e., 6.02), from 5.5 to 6.5 in steps of 10^{-4}, confirming that the MSE reported in Table 3 corresponds to a wide minimum. We further tried the same function, changing the input data from $d_j(t-1)$ to $d_j(t-2)$ and $d_j(t-3)$, thus artificially changing the time lag of the relationship; the obtained MSEs increase to respectively 228,174 and 231,245, i.e., close to the null model MSE, confirming the relevance of this time lag. We further checked whether a similar relation can be obtained when only basic operators (i.e., sum and multiplication) are allowed. Notably, the relation between the two airports is lost, and the dependence of LSZH with its past seen in Table 2 is recovered.

Reviewing the results reported in Table 3, it can be appreciated that several relationships connect the airports of London Heathrow, Roma Fiumicino, and Zurich. We specifically identified the following dependencies: LSZH → EGLL, EGLL → LIRF, and LSZH → LIRF. A final question that could be posed is whether multivariate relationships also exist, i.e., relationships involving these three airports at the same time. This can easily be checked by finding models predicting the dynamic of one airport, using both its own past, and the past of the two other airports. Results are reported in Table 4; note that, in order to bound the computational cost, only data for the two previous time steps (i.e., $t-1$ and $t-2$) have been included. No major changes appear, compared to what we previously reported in Table 3; both EGLL and LIRF dynamics seem to be driven by LSZH. On the other hand, the relationship EGLL → LIRF does not appear. This may indicate that it was a spurious relationship, due to LSZH acting on both airports at the same time, but also that the SR may fail to recover complex relations involving multiple variables.

Table 4. Analysis of the relationships between EGLL, LIRF, and LSZH. The second and third columns report the MSEs corresponding to the data and to the null model.

Equation	MSE	rMSE
$d_{EGLL}(t) = d_{EGLL}(t-1)erf(d_{LSZH}(t-1)^2)$	903,928	923,596
$d_{LIRF}(t) = 1.06 d_{LIRF}(t-1)\cos(erf(d_{LSZH}(t-1)))$	292,575	299,627
$d_{LSZH}(t) = 202.70 + 0.31 d_{LSZH}(t-1)$	203,428	203,428

5. Discussion and Conclusions

In this contribution we have explored some possibilities offered by symbolic regression (SR) in the context of the study of the dynamics of air transport delays. While SR is a family of techniques that have found wide application in science, this is the first time, to the best of our knowledge, that it has been applied to this specific problem. It is now time to try to answer the question we initially posed: is SR a viable alternative to more classical methodologies?

As seen throughout Sections 3 and 4, SR yields an important added value: the possibility of recovering the functional form of the relationship between two sets of data. This can involve estimating how the delay at one airport can be forecasted from its own past (Section 3), or from the past of one or more other airports (Section 4), akin to a predictive causality analysis. A researcher interested in obtaining this type of information would usually have to choose between two alternative approaches. On the one hand, they can test specific closed-form relationships (e.g., the presence of a statistically significant linear correlation), usually ensuring high sensitivity but limited flexibility. On the other hand, they can resort to more general data analysis models, the best example being Deep Learning; while these can in principle detect the presence of any type of relationship, their black-box nature limits the quantity of information that can be made explicit. SR seems to offer the best of both worlds: the possibility of detecting complex relationships between data, while presenting the results in a clear and easily interpretable way.

Following the no-free-lunch theorem, these advantages come at an important cost, here a computational one. SR involves the exploration of a potentially infinite space of functions; the use of evolutionary optimisation algorithms, as included in the PySR package, helps with, but does not completely solve, the problem. To illustrate, Figure 7 reports the evolution of the computational cost for retrieving the relationships for London Heathrow, as a function of the maximum complexity and the number of iterations in the genetic optimisation. Results are in the order of hours, compared to the fractions of a second required to evaluate a Granger Causality. In addition, the stochastic nature of the exploration of such space implies that stable results are only obtained after a large number of iterations (see Figure 2). Note that this computational cost becomes especially relevant when the objective is the study of delay propagation instances in large networks. As the cost grows as N^2, where N is the number of considered airports, an analysis of the network composed of the top-100 airports in a region may take more than one year in a standard computer—again, compared to the minutes of the Granger Causality alternative. This also hinders any real-time application of this approach. All in all, while SR can, for instance, make explicit relationships between observables at different airports in a way far richer than a Granger Causality test (see Table 3), the associated computational cost may become a major barrier.

As a final point worth discussing, SR has here been used in an isolated fashion; still, it can also be integrated with complementary machine and Deep Learning techniques, offering some interesting opportunities. To illustrate, one of these could be neuro-symbolic artificial intelligence, i.e., a hybrid approach to data analysis combining classical neural networks with symbolic representation of knowledge [56,57]; interpretable symbolic equations could be derived from learned neural representations, with neural networks thus acting as a simplification layer [58]. Reversing the logic, SR can also be used to replace individual elements inside a neural network, or full sub-networks [59], something known as neural network compression [60].

https://doi.org/10.3390/aerospace13060535

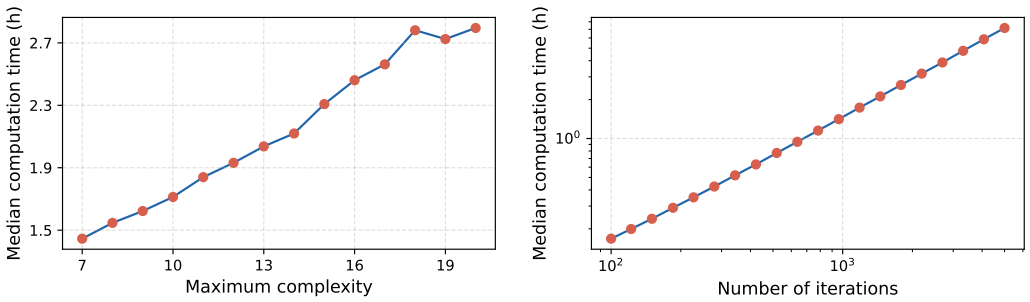


Figure 7. Evolution of the computational cost of extracting the relationship between past and future delays for London Heathrow, EGLL, i.e., as in Table 2. **Left** and **right** panels respectively correspond to results as a function of the maximum complexity (i.e., number of operators being combined) and of the number of iterations. Results correspond to the median of 50 independent realisations; computations were performed on a single core of an AMD 7282 processor at 2.8 GHz.

Whether using only SR, or a combination of this with other machine and Deep Learning techniques, the outcome could be models that are both predictive and explainable, and hence of high usefulness in the context of air transport. This is especially relevant in the context of flight delays, as these propagate between connecting flights, leading to large disruptions in the overall schedule. The aforementioned models could allow airlines and airport operators to proactively manage disruptions. To illustrate, airlines can optimise flight schedules and add buffers where delays are predicted to occur; in turn, airport authorities can use these models to efficiently manage ground operations, gate assignment, and capacity planning. Ultimately, airports and airline managers could take necessary measures to minimise the losses caused by delays and increase the efficiency of the aviation system, thus improving the mobility of passengers and of society at large.

Funding: This project has received funding from Grant CNS2023-144775 funded by MICIU/AEI/10.13039/501100011033 by "European Union NextGenerationEU/PRTR". This work was partially supported by the María de Maeztu project CEX2021-001164-M funded by the MICIU/AEI/10.13039/501100011033 and FEDER, EU.

Data Availability Statement: The data presented in this study are available in EUROCONTROL's Aviation Data Repository for Research, at https://www.eurocontrol.int/dashboard/aviation-data-research (accessed on 1 October 2025).

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following ICAO airport codes are used in this manuscript:

EGLL	London Heathrow
LFPG	Paris Charles de Gaulle
EHAM	Amsterdam Schiphol
EDDF	Frankfurt
LEMD	Madrid Barajas
LEBL	Barcelona El Prat International
EDDM	Munich
EGKK	London Gatwick
LIRF	Rome Fiumicino
LFPO	Paris Orly
EIDW	Dublin

LSZH	Zurich
EKCH	Copenhagen
LEPA	Palma De Mallorca
LPPT	Lisboa
ENGM	Oslo Gardermoen
EGCC	Manchester
EGSS	London Stansted
LOWW	Vienna International Airport
ESSA	Stockholm Arlanda

References

1. Cook, A.; Tanner, G.; Williams, V.; Meise, G. Dynamic cost indexing–Managing airline delay costs. *J. Air Transp. Manag.* **2009**, *15*, 26–35. [CrossRef]
2. Britto, R.; Dresner, M.; Voltes, A. The impact of flight delays on passenger demand and societal welfare. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 460–469. [CrossRef]
3. Wang, Y.; Cao, Y.; Zhu, C.; Wu, F.; Hu, M.; Duong, V.; Watkins, M.; Barzel, B.; Stanley, H.E. Universal patterns in passenger flight departure delays. *Sci. Rep.* **2020**, *10*, 6890. [CrossRef] [PubMed]
4. Mitsokapas, E.; Schäfer, B.; Harris, R.J.; Beck, C. Statistical characterization of airplane delays. *Sci. Rep.* **2021**, *11*, 7855. [CrossRef]
5. Sun, L.L.; Hu, Y.P.; Zhu, C.P. Scaling invariance in domestic passenger flight delays in the United States. *Phys. A Stat. Mech. Its Appl.* **2023**, *611*, 128445. [CrossRef]
6. Wu, Y.; Wu, J.; Zhang, Y.; Cui, Y.; Zhang, Q.; Sun, L. Universality, criticality and complexity of flight delay propagation. *Phys. A Stat. Mech. Its Appl.* **2025**, *683* , 131209. [CrossRef]
7. Olivares, F.; Sun, X.; Wandelt, S.; Zanin, M. Measuring landing independence and interactions using statistical physics. *Transp. Res. Part E Logist. Transp. Rev.* **2023**, *170*, 102998. [CrossRef]
8. Olivares, F.; Zanin, M. Beyond Classical Models: Statistical Physics Tools for the Analysis of Time Series in Modern Air Transport. *arXiv* **2025**, arXiv:2507.20927. [CrossRef]
9. Tang, Z.; Huang, S.; Han, S. Recent progress about flight delay under complex network. *Complexity* **2021**, *2021*, 5513093. [CrossRef]
10. Cai, Q.; Alam, S.; Duong, V.N. A spatial–temporal network perspective for the propagation dynamics of air traffic delays. *Engineering* **2021**, *7*, 452–464. [CrossRef]
11. Gurtner, G.; Delgado, L.; Valput, D. An agent-based model for air transportation to capture network effects in assessing delay management mechanisms. *Transp. Res. Part C Emerg. Technol.* **2021**, *133*, 103358. [CrossRef]
12. Gil-Rodrigo, S.; Zanin, M. Low Cost Carriers induce specific and identifiable delay propagation patterns: An analysis of the EU and US systems. *IEEE Access* **2024**, *12*, 75323–75336. [CrossRef]
13. Zanin, M. Reconstructing functional networks of air transport delay propagations with minimal information. *Phys. A Stat. Mech. Its Appl.* **2025**, *659*, 130318. [CrossRef]
14. Zanin, M.; Belkoura, S.; Zhu, Y. Network analysis of Chinese air transport delay propagation. *Chin. J. Aeronaut.* **2017**, *30*, 491–499. [CrossRef]
15. Mazzarisi, P.; Zaoli, S.; Lillo, F.; Delgado, L.; Gurtner, G. New centrality and causality metrics assessing air traffic network interactions. *J. Air Transp. Manag.* **2020**, *85*, 101801. [CrossRef]
16. Truong, D. Using causal machine learning for predicting the risk of flight delays in air transportation. *J. Air Transp. Manag.* **2021**, *91*, 101993. [CrossRef]
17. Jia, Z.; Cai, X.; Hu, Y.; Ji, J.; Jiao, Z. Delay propagation network in air transport systems based on refined nonlinear Granger causality. *Transp. B Transp. Dyn.* **2022**, *10*, 586–598. [CrossRef]
18. Pastorino, L.; Zanin, M. Air delay propagation patterns in Europe from 2015 to 2018: An information processing perspective. *J. Phys. Complex.* **2022**, *3*, 015001. [CrossRef]
19. Zhu, Q.; Chen, S.; Guo, T.; Lv, Y.; Du, W. A spatio-temporal approach with self-corrective causal inference for flight delay prediction. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 20820–20831. [CrossRef]
20. Yu, B.; Guo, Z.; Asian, S.; Wang, H.; Chen, G. Flight delay prediction for commercial air transport: A deep learning approach. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *125*, 203–221. [CrossRef]
21. Gui, G.; Liu, F.; Sun, J.; Yang, J.; Zhou, Z.; Zhao, D. Flight delay prediction based on aviation big data and machine learning. *IEEE Trans. Veh. Technol.* **2019**, *69*, 140–150. [CrossRef]
22. Tan, X.; Liu, Y.; Liu, D.; Zhu, D.; Zeng, W.; Wang, H. An attention-based deep convolution network for mining airport delay propagation causality. *Appl. Sci.* **2022**, *12*, 10433. [CrossRef]

23. Qu, J.; Wu, S.; Zhang, J. Flight delay propagation prediction based on deep learning. *Mathematics* **2023**, *11*, 494. [CrossRef]
24. Zhong, Q.; Yu, Y.; Huang, Y.; Zhang, T. Prediction and optimization of civil aviation flight delays based on machine learning algorithms. *Int. J. Comput. Intell. Syst.* **2025**, *18*, 189. [CrossRef]
25. Langley, P. Data-driven discovery of physical laws. *Cogn. Sci.* **1981**, *5*, 31–54. [CrossRef]
26. Langley, P. *Scientific Discovery: Computational Explorations of the Creative Processes*; MIT Press: Cambridge, MA, USA, 1987.
27. Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [CrossRef]
28. Schmidt, M.; Lipson, H. Distilling free-form natural laws from experimental data. *Science* **2009**, *324*, 81–85. [CrossRef]
29. La Cava, W.; Burlacu, B.; Virgolin, M.; Kommenda, M.; Orzechowski, P.; de França, F.O.; Jin, Y.; Moore, J.H. Contemporary symbolic regression methods and their relative performance. *Adv. Neural Inf. Process. Syst.* **2021**, *2021*, 1.
30. Makke, N.; Chawla, S. Interpretable scientific discovery with symbolic regression: A review. *Artif. Intell. Rev.* **2024**, *57*, 2. [CrossRef]
31. Kronberger, G.; Burlacu, B.; Kommenda, M.; Winkler, S.M.; Affenzeller, M. *Symbolic Regression*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2024.
32. Wu, C.; Zhang, Y. Enhancing the shear-stress-transport turbulence model with symbolic regression: A generalizable and interpretable data-driven approach. *Phys. Rev. Fluids* **2023**, *8*, 084604. [CrossRef]
33. Song, H.; Ma, M.; Yi, C.; Shao, Z.; Bai, R.; Yan, C. An innovative modification to the Menter shear-stress transport turbulence model employing the symbolic regression approach. *Phys. Fluids* **2024**, *36*, 065108 . [CrossRef]
34. Ding, D.; Wang, Q.; Chen, Q.; He, L. Symbolic Regression-Based Modeling for Aerodynamic Ground-to-Flight Deviation Laws of Aerospace Vehicles. *Aerospace* **2025**, *12*, 455. [CrossRef]
35. Wu, C.; Zhang, S.; Zhang, Y. Development of a generalizable data-driven turbulence model: Conditioned field inversion and symbolic regression. *AIAA J.* **2025**, *63*, 687–706. [CrossRef]
36. Kefalas, M.; Santiago Rojo, J.d., Jr.; Apostolidis, A.; Van Den Herik, D.; van Stein, B.; Bäck, T. Explainable artificial intelligence for exhaust gas temperature of turbofan engines. *J. Aerosp. Inf. Syst.* **2022**, *19*, 447–454. [CrossRef]
37. Sun, Y.P.; Feng, H.; Zheng, B.; Wen, J.R.; Chao, A.F.; Fei, C.W. Multi-Agent Reinforcement Symbolic Regression for the Fatigue Life Prediction of Aircraft Landing Gear. *Aerospace* **2025**, *12*, 718. [CrossRef]
38. Brandejsky, T. The Train Delay Model Developed by the Genetic Programming Algorithm. *J. Adv. Transp.* **2022**, *2022*, 8858756. [CrossRef]
39. Diebold, F.X. *Elements of Forecasting*; South-Western College Pub.: Cincinnati, OH, USA, 1998.
40. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [CrossRef]
41. Granger, C.W. Some recent development in a concept of causality. *J. Econom.* **1988**, *39*, 199–211. [CrossRef]
42. Acharya, K.; Zanin, M. Information decomposition of airport dynamics: A study of Europe and US. *Transp. Res. Interdiscip. Perspect.* **2025**, *31*, 101466. [CrossRef]
43. Olivares, F.; Zanin, M. Quantifying deviations from Gaussianity with application to flight delay distributions. *Entropy* **2025**, *27*, 354. [CrossRef]
44. Du, W.; Zhu, Q.; Chen, S.; Guo, T.; Zhu, Y. Flight Delay Prediction via Learning Long Short-Term Relationship Between Airports. *IEEE Intell. Transp. Syst. Mag.* **2024**, *17*, 6–19. [CrossRef]
45. Wandelt, S.; Chen, X.; Sun, X. Networks for flight delay analysis: A scoping review and research agenda. *IEEE Trans. Netw. Sci. Eng.* **2025**, *12*, 1250–1266. [CrossRef]
46. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]
47. Tian, Y.; Zhou, W.; Viscione, M.; Dong, H.; Kammer, D.S.; Fink, O. Interactive symbolic regression with co-design mechanism through offline reinforcement learning. *Nat. Commun.* **2025**, *16*, 3930. [CrossRef]
48. Guimerà, R.; Sales-Pardo, M. Bayesian symbolic regression: Automated equation discovery from a physicist's perspective. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2026**, *384*, 20250089. [CrossRef]
49. Kamienny, P.A.; Lample, G.; Lamprier, S.; Virgolin, M. Deep generative symbolic regression with Monte-Carlo-tree-search. In *Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023*; Curran Associates, Inc.: Red Hook, NY, USA, 2023; pp. 15655–15668.
50. Cranmer, M. Interpretable machine learning for science with PySR and SymbolicRegression.jl. *arXiv* **2023**, arXiv:2305.01582.
51. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer: Berlin/Heidelberg, Germany, 2006.
52. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 63623. [CrossRef]
53. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
54. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
55. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]

56. Sarker, M.K.; Zhou, L.; Eberhart, A.; Hitzler, P. Neuro-symbolic artificial intelligence: Current trends. *AI Commun.* **2022**, *34*, 197–209. [CrossRef]
57. Bhuyan, B.P.; Ramdane-Cherif, A.; Tomar, R.; Singh, T. Neuro-symbolic artificial intelligence: A survey. *Neural Comput. Appl.* **2024**, *36*, 12809–12844. [CrossRef]
58. Cranmer, M.; Sanchez Gonzalez, A.; Battaglia, P.; Xu, R.; Cranmer, K.; Spergel, D.; Ho, S. Discovering symbolic models from deep learning with inductive biases. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17429–17442.
59. Wei, W.; Lu, Q.; Huang, C.; Luo, J. Refining Neural Network with Symbolic Regression. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, Malaga, Spain, 14–18 July 2025; pp. 923–926.
60. Neill, J.O. An overview of neural network compression. *arXiv* **2020**, arXiv:2006.03669. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.