A Type-2 Fuzzy Modelling Framework for Aircraft Taxi-Time Prediction

Olusayo Obajemu, Mahdi Mahfouf and David Allerton
Department of Automatic Control & Systems Engineering
The University of Sheffield, Sheffield S1 3JD, U.K.
Email: o.obajemu@sheffield.ac.uk; m.mahfouf@sheffield.ac.uk; d.j.allerton@sheffield.ac.uk

Michal Weiszer and Jun Chen
School of Engineering
University of Lincoln, Lincoln LN6 7TS, U.K.
Email: mweiszer@lincoln.ac.uk; juchen@lincoln.ac.uk

***Abstract*—Knowing aircraft taxiing times precisely a-priori is increasingly important for any airport management system. This work presents a new approach to estimating and characterising the taxi-time of aircraft based on historical information. The approach makes use of the interval type-2 fuzzy logic system, which provides more robustness and accuracy than the conventional type-1 fuzzy system. To compensate for erroneous modelling assumptions, the error distribution of the model is further analysed and an error compensation strategy is developed. Results, when tested on a real dataset for Manchester Airport, show improved taxi-time accuracy and generalisation capability over a wide range of modelling assumptions when compared with existing fuzzy systems and linear regression-based methods.**

***Index Terms*—taxi, fuzzy, uncertainty, aircraft, airport operations, ground movement.**

I. Introduction

The advent of automated air traffic management systems at airports over the last decade has prompted and increased the need for accurate taxiing time predictions. Airport ground movements contributes significantly to the annual average delays experienced at major airports [1], [2]. This is because the ground movement serves as an important link between other aspects of ground operations such as departure sequencing and gate allocation[1] [3], [4]. Sophisticated routing and scheduling algorithms are being developed and deployed in order to address inefficiencies of airport ground movements and these algorithms need a-priori specification of accurate taxiing times [4], [5]. In [6]–[9], taxiing times were used to schedule aircraft departures as well as stand holds allocation which underscores the importance of accurate prediction of taxiing times. Furthermore, knowing taxi times in advance can provide very valuable information to scheduling algorithms to mitigate the effects of predicted delays.

Until very recently, the common method for taxiing time estimation focused on the use of the mean times for source-destination pairs. In order to account for the variables/factors that influence taxi times such as the size of an aircraft, machine learning approaches are increasingly gaining popularity. A literature survey of implemented methods reveals that there is a concentration on linear regression-based methods which is to be expected as these methods tend to be simple to implement and interpret [3]. This easy interpretation is in contrast to the so-called black modelling approaches (such as neural networks) which, apart from being opaque in interpretability, can be computationally expensive. Fuzzy logic provides an equivalent (or better) non-linear mapping capability when compared to these black-box methods. They also have the advantage of being able to represent complex statements via linguistic statements which makes them easily accessible thereby facilitating the knowledge fusion with expert information. To take advantage of these attractive properties of fuzzy reasoning, [10] have used the Mamdani-type fuzzy logic system to predict the taxi time, which shows an average increase in prediction accuracy of approximately 10% over the linear regression approaches with which results were compared. However, in this model, the effect of uncertainty in predictions were not accurately captured since the membership functions (explained in section 3) are represented by crisp values which limit the handling of the so-called linguistic uncertainties[2]. Type-2 fuzzy systems enable these uncertainties to be adequately handled.

It is widely acknowledged that the taxi-time prediction problem is fraught with uncertainties [12]. These uncertainties come in different forms including the continuously changing airport environment and the different types of undercarriage assembly [13]. The adverse effects of uncertainties in prediction problems is well documented in [14]. In the taxiing problem, uncertainty reduces predictability and inadequate handling of these uncertainties can lead to the sophisticated algorithms (which rely on accurate taxiing time prediction) becoming ineffective. In this work, a hierarchical framework where the taxi time model is developed using a type-2 fuzzy logic system (T2FLS) to allow for handling linguistic uncertainties is proposed. The second stage is an error compensation scheme similar to that developed in our earlier work in [15], which allows for uncertainty handling. To the best of our knowledge, this is the first time that such an approach has been used to address the above challenges implicit in the taxiing time prediction problem. The remainder of the paper is organised as follows: Section 2 presents the data used in

[1]reducing ground delay at the gate thereby saving fuel and reducing adverse environmental effects.

[2]linguistic uncertainty represents a scenario the meaning of a word is not fixed

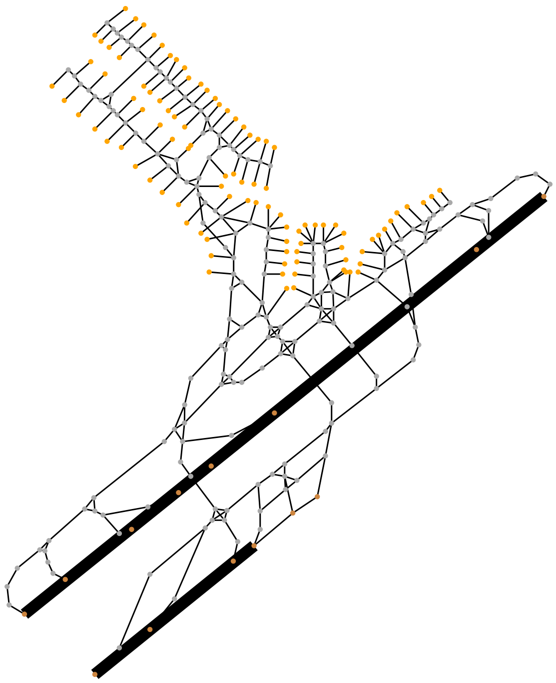


Fig. 1: Layout of Manchester Airport. Figure generated from publicly available data (openstreetmap.org) [11].

this work. Section 3 briefly describes the proposed modelling framework, which includes the type-2 fuzzy logic system and the proposed error compensation strategy. Section 4 analyses and discusses the experimental results from the study as well as a comparative study of the proposed approach with other popular algorithms. Finally, Section 5 concludes with a summary of the main contributions of this research and recommendations for future research.

II. The Data

Historical data from Manchester Airport (Fig. 1) have been used to develop the aircraft taxi time model based on the proposed framework (discussed in Section III)[3]. A total of 1413 data points were observed which contains information relating to the factors that affect the output variable (the taxi-time). A more rigorous statistical analysis has been performed in [5] and this has identified the 15 variables most significant for predicting the taxi times of aircraft. A smaller sample of these variables are listed in Table 1.

III. Modelling Framework

The modelling framework involves two stages; the first stage consists of developing a type-2 fuzzy model. The second stage involves analysing the errors in the model in order to produce an error compensation strategy. Analysis of these errors involves fitting mixtures of a multi–dimensional Gaussian distribution over the inputs and the error are as described in Section IIIB.

[3]Details of the data gathering process are discussed in [12]

TABLE I: The input variables included for study in this research.

Variable	Meaning
Distance (metres)	The total taxiing distance
Arr/Dep	A binary variable which indicates if an aircraft is arriving or departing
Size	Another binary variable. 0 represents a large aircraft and 1 represents a small aircraft
Q	Q values represent the number of aircraft that stop taxiing during the time when the aircraft is taxiing
N	The total number of taxiing aircraft
Operational Mode	Indicates the mode of operation of the airport at the time a the aircraft starts taxiing

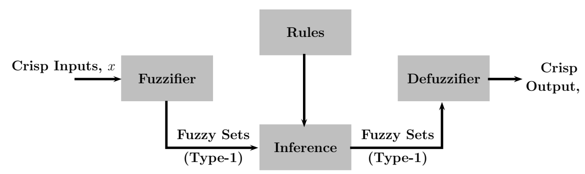


Fig. 2: Type-1 Fuzzy Logic System.

A. A Type-2 Fuzzy Model

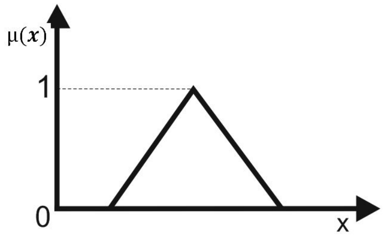
This section provides the ‘rationale for using T2FLSs as a better alternative for the T1FLS. As already mentioned in section 1, fuzzy logic has the advantage of representing complex systems using human intuition. The block diagram of a conventional fuzzy logic system is shown in Fig. 2.

It can be seen that a T1FLS represents a mapping from an input space (X) to an output space (Y). As discussed in [3], the simple block represented by Fig. 2 can represent any non-linear function [16], [17]. The ability to incorporate human-like information/reasoning is mainly due to the fuzzy sets embedded in the rules section of the block diagram. Given a FLS with n inputs ($x_1 \in X_1, x_2 \in X_2, \cdots, x_n \in X_n$) and one output ($y \in Y$) and a rule-base composing of c-rules, the ith rule R_i can be expressed as follows:

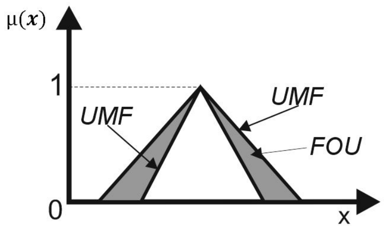
$$R_i : \text{IF } x_1 \text{ is } A_1^i \text{ and } x_2 \; is \; A_2^i \; \cdots \text{ and } x_n \; is \; A_n^i \text{ THEN} \\ y_i \text{ is } h_i(\mathbf{x}) \;\; \mathbf{x} \in R^n \quad (1)$$

A_j^i and h_i represent the jth membership function (MF) and the consequent of the ith rule respectively for $j = 1, 2, \cdots, n$ and $i = 1, 2, \cdots, c$. A_j^i is a fuzzy set (Fig. 3) which is a mathematical representation of the subjective linguistic knowledge. Further details about the meaning of fuzzy sets are given in [3].

In order to be able to handle the so-called linguistic uncertainties (see [3] and [18] for details), type-2 fuzzy sets were



(a) Type-1 Fuzzy Set.



(b) Type-2 Fuzzy Set.

Fig. 3: Fuzzy Sets.

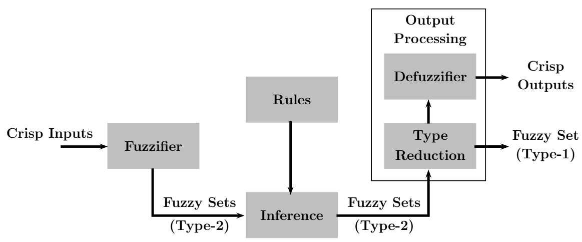


Fig. 4: Type-2 Fuzzy Logic System.

introduced (Fig . 3b). Type-2 fuzzy sets allow for representing the membership functions with another fuzzy set (called a secondary membership function) which means that the MF is no longer a crisp value as in the case of the type-1 fuzzy set. Any fuzzy logic system that utilises at least one type-2 fuzzy set in its rules is called a type-2 fuzzy logic system. Compared with a T1FLS, the T2FLS has an additional block in its block diagram as shown in Fig. 4 owing to the need to reduce the type-2 fuzzy sets after the inference procedure to a type-1 fuzzy set. It is worth noting that the interval type-2 fuzzy logic system was used in this study [3].

As in the case of the T1FLS, the T2FLS is also a mapping from the input domain, X to the output domain, Y and can be represented by the following:

$$\hat{y} = \left(\sum_{i=1}^{c} h_i \overline{f}_i + \sum_{i=1}^{c} h_i \underline{f}_i \right) \Big/ \left(\sum_{i=1}^{c} \overline{f}_i + \sum_{i=1}^{c} \underline{f}_i \right) \tag{2}$$

where $\hat{y}$ is the output of the FLS and $\overline{f}_i$ and $\underline{f}_i$ represent the upper and lower membership functions respectively, which depends on the input, X.

T2FLSs have been shown to be more robust to noise and uncertainties due to the extra degree of freedom in their membership functions [18]. Introducing robustness and improved accuracy in the taxi–time prediction through using these T2FLSs can only be beneficial. When there is limited expert information, fuzzy logic systems can be identified using automatic rule generation methods such as those proposed in [3], [15], [18]. Usually, the parameters of the fuzzy logic system are tuned to minimise an objective function using the root mean square error (RMSE) where the RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{3}$$

where N is the number of data points, y_i and $\hat{y}_i$ are the ith actual output and the corresponding predicted output of the model respectively. The RMSE works well for homoscedastic data [19]. However, it can sometimes be difficult to validate the homoscedastic assumption which has motivated the error compensation strategy discussed in the next section.

B. Uncertainty Modelling and the Error Compensation Strategy

The second stage of the modelling framework involves including a Gaussian Mixture Modelling (GMM)-based error compensation strategy, which has been discussed in our earlier work in [15]. After the fuzzy modelling stage, for each sets of input variables, the residual errors can be obtained. The GMM model is then fitted into the combination of the input variables and the residuals. The error compensation strategy serves two purposes: 1. To provide distribution of the error for a given input which can be used to represent the confidence band in predictions. 2. To compensate for biases due to unverifiable modelling assumptions at the beginning of the modelling process. The steps for constructing the error compensation scheme are described as follows:

1) Arrange the input variables and the error using the following equation:

$$X_E = [XE] \tag{4}$$

where E is the corresponding error which is the difference between predicted output and the measured output for the corresponding input X.

2) Initialise the GMM parameters using randomly chosen values for a predefined number of Gaussian components, K. The parameters to be initialised include the mixing coefficients π_k, the mean μ_k and the covariance matrix σ_k. The GMM is defined by the following equation:

$$P(x_n^e|\pi,\mu,\sigma) = \sum_{k}^{K} \pi_k g(x_n^e|\mu_k,\sigma_k) \tag{5}$$

where $P(x_n^e|\pi,\mu,\sigma)$ is the unconditional probability of a particular data point x_n^e, π_k is the mixing coefficient for the kth Gaussian with mean μ_k and covariance matrix σ_k. $g(x_n^e|\mu_k,\sigma_k)$ is the probability of that particular data point x_n^e given that it belongs to the kth Gaussian component. It should be noted that $\sum_k^K \pi_k = 1$.

3) Calculate the membership weight for the nth data point x_e^n for a particular cluster k using the following equation:

$$z_k(x_n^e) = \frac{g(x_n^e|\mu_k, \sigma_k)}{\sum_k^K \pi_k g(x_n^e|\mu_k, \sigma_k)} \tag{6}$$

This is called the E-Step of the expectation maximisation algorithm.

4) Calculate new parameter values given by the sets of equations below:

$$\pi_k = \frac{1}{N}\sum_{n=1}^{N} z_k(x_n^e) \tag{7}$$

$$\mu_k = \frac{\sum_{n=1}^{N} z_k(x_n^e)x_n^e}{\sum_{n=1}^{N} z_k(x_n^e)} \tag{8}$$

$$\sigma_k = \frac{\sum_{n=1}^{N} z_k(x_n^e - \mu_k)(x_n^e - \mu_k)^T}{\sum_{n=1}^{N} z_k(x_n^e)} \tag{9}$$

This is called the M-Step of the expectation maximisation algorithm.

5) Steps 1-4 above are repeated until convergence. It is worth noting that convergence of this algorithm can be aided significantly by pre-processing the data using clustering algorithms such as the k-means algorithm. Convergence can be detected when there are no further changes to the values of the parameters or by computing the log-likelihood (under independence and identically distributed assumption). The number of components is selected using the Bayesian Information criterion discussed in [15].

6) For a given input x_i, the distribution of the error is calculated using Bayes rule as given by the following equation:

$$P(e|x_i) = \frac{P(x_i, e)}{P(x_i)} \tag{10}$$

$$= \frac{P(x_i, e)}{\int P(x_i, e)de} \tag{11}$$

7) The expectation of the error is given by the following equation:

$$M_e(x_i) = \int eP(x|e_i)de \tag{12}$$

In addition to the above, the standard deviation, which is used to construct the confidence interval in the predictions, is given by the following equation:

$$S_e(x_i) = \sqrt{\int (e - M_e)^2 P(e|x_i)de} \tag{13}$$

8) The error in the predictions is then compensated by using the following equation:

$$y_i^c = y_i - M_e(x_i) \tag{14}$$

The error compensation scheme, as defined in step 8 above, is based on the assertion that when $M_e(x_i)$ is negative, the predictions are negatively biased. The error compensation strategy then compensates for this bias.

TABLE II: Comparison of proposed approach with popular techniques from the literature. LR means Linear Regression, M FRBS means the Mamdani Fuzzy Rule-Based System. Proposed* is the compensated model.

Method	**$\pm$2 mins**	**$\pm$3 mins**
LR	70.11%)	84.6%
M FRBS [10]	88.4%	93.7%
Proposed	90.01%	95.21%
Proposed*	91.32%	97.54%

IV. Results

The proposed approach was tested on the experimental dataset described in Section 2. This dataset from Manchester Airport contains 1413 data points with 15 input variables and 1 output variable. The dataset was divided into two parts: 69% for training the model as well as fitting the GMM model and 31% for testing the generalising capability of such a prediction model. The steps discussed in section 4 were implemented to train an interval type-2 fuzzy model with an error compensation strategy. The performance criterion of choice for this study is the RMSE error defined in Eqn. 3.

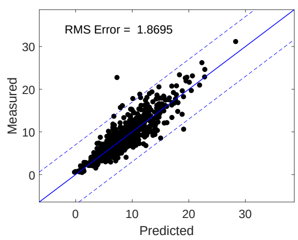
A. Fuzzy model result without error compensation

The results of the fuzzy model without error compensation is shown in Fig. 5a and 5b for the training and testing data respectively. Compared with linear regression results shown in Fig. 5c and Fig. 5d (for training and testing data respectively), the fuzzy model shows an increase in performance of approximately 15%.

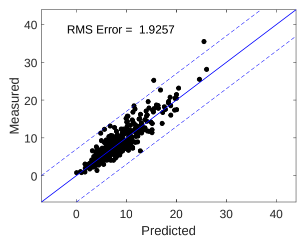
Fig. 6 shows the error distribution of 10 data points. As can be seen from the figure, the uncertainty in predictions tend to increase with increasing taxi-times which is reasonable as the longer it takes an airplane to taxi, the more the tendency for the time it takes to taxi to-and-from gates to be uncertain. Table 2 shows that, the proposed approach provides a competitive advantage over popular techniques for taxi-time prediction. Fig. 7 shows the three-dimensional plot of size of aircraft and taxi distance in metres against the predicted taxi time in minutes. Larger aircraft on the average take longer to taxi.

V. Conclusion

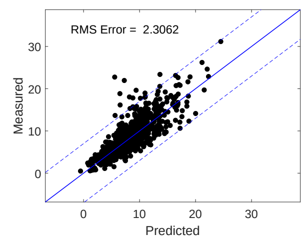
This study has presented a new approach to estimating the taxi-times of aircraft. The approach is based on the interval type-2 fuzzy logic systems which provides more robustness than the conventional type-1 fuzzy logic system. The advantage of fuzzy systems modelling over other types of non-linear mapping functions is the ease of interpretability and the ability to incorporate linguistic information/expert knowledge in a seamless manner. The second stage of the modelling process includes fitting a GMM model over the input variables and the error from the predictions. This model can both serve as an error compensation scheme as well provide a confidence band in the predictions. Results, when compared with existing



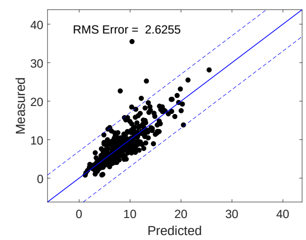
(a) Fuzzy Model Results Before Error Compensation (Training Data).



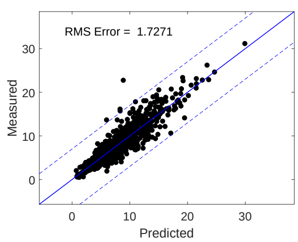
(b) Fuzzy Model Results before Error Compensation (Testing Data).



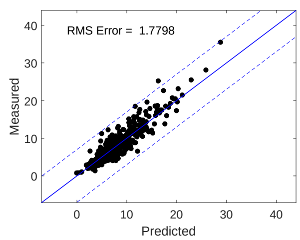
(c) Linear Regression Results before Error Compensation (Training Data).



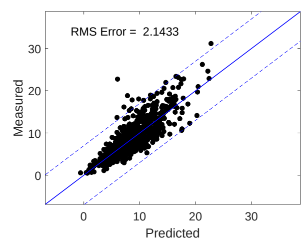
(d) Linear Regression Results before Error Compensation (Testing Data).



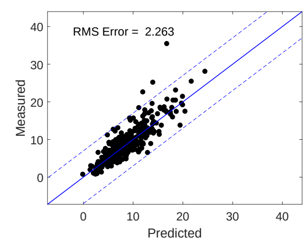
(e) Fuzzy Model Results after Error Compensation (Training Data).



(f) Fuzzy Model Results after Error Compensation (Testing Data).



(g) Linear Regression Results after Error Compensation (Training Data).



(h) Linear Regression Results after Error Compensation (Training Data).

Fig. 5: Results.

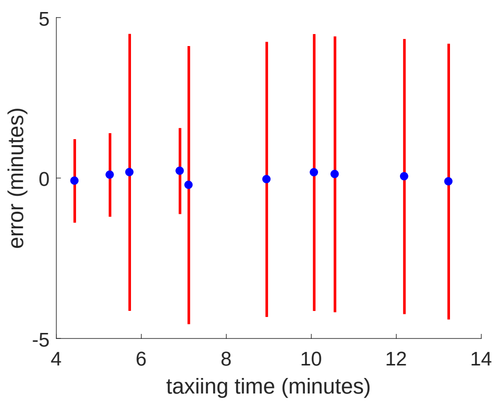


Fig. 6: Error Distribution of 10 of the data points.

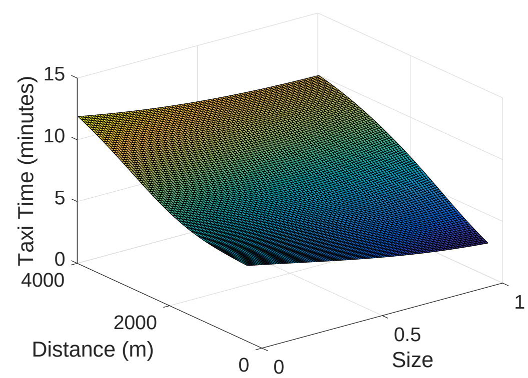


Fig. 7: Three-dimensional plot of two inputs (taxiing distance in metres and size of aircraft [0 for large, 1 for small]) against the predicted taxi-time (minutes).

algorithms show the efficacy of the proposed approach. The general type-2 fuzzy system will be considered in future work.

References

[1] J. Chen and P. Stewart, "Planning aircraft taxiing trajectories via a multi-ojective immune optimisation," in *Natural Computation (ICNC), 2011 Seventh International Conference on*, vol. 4. IEEE, 2011, pp. 2235–2240.

[2] A. G. Marin, "Airport management: taxi planning," *Annals of Operations Research*, vol. 143, no. 1, pp. 191–202, 2006.

[3] J. M. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall PTR Upper Saddle River, 2001.

[4] J. A. Atkin, E. K. Burke, and S. Ravizza, "The airport ground movement problem: Past and current research and future directions," in *Proceedings of the 4th International Conference on Research in Air Transportation (ICRAT), Budapest, Hungary*, 2010, pp. 131–138.

[5] S. Ravizza, J. A. Atkin, M. H. Maathuis, and E. K. Burke, "A combined statistical approach and ground movement model for improving taxi time estimations at airports," *Journal of the Operational Research Society*, vol. 64, no. 9, pp. 1347–1360, 2013.

[6] J. A. Atkin, E. K. Burke, J. S. Greenwood, and D. Reeson, "Hybrid metaheuristics to aid runway scheduling at london heathrow airport," *Transportation Science*, vol. 41, no. 1, pp. 90–106, 2007.

[7] J. A. Atkin, E. K. Burke, and J. S. Greenwood, "Tsat allocation at london heathrow: the relationship between slot compliance, throughput and equity," *Public Transport*, vol. 2, no. 3, pp. 173–198, 2010.

[8] P. Burgain, E. Feron, and J.-P. Clarke, "Collaborative virtual queue: Benefit analysis of a collaborative decision making concept applied to congested airport departure operations," *Air Traffic Control Quarterly*, vol. 17, no. 2, pp. 195–222, 2009.

[9] "Airport cdm implementation manual. technical report. version 3.1," http://www.euro-cdm.org/ library/cdm_implementation_manual.pdf, ac-

cessed: 2016-09-30.
[10] J. Chen, S. Ravizza, J. A. Atkin, and P. Stewart, "On the utilisation of fuzzy rule-based systems for taxi time estimations at airports," in *OASIcs-OpenAccess Series in Informatics*, vol. 20. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2011.
[11] J. Chen, M. Weiszer, G. Locatelli, S. Ravizza, J. A. Atkin, P. Stewart, and E. K. Burke, "Toward a more realistic, cost-effective, and greener ground movement through active routing: A multiobjective shortest path approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3524–3540, 2016.
[12] A. Brownlee, J. Atkin, J. Woodward, U. Benlic, and E. Burke, "Airport ground movement: real world data sets and approaches to handling uncertainty," *Proc. of the Practice and Theory of Automated Timetabling, York, UK*, 2014.
[13] D. Allerton, *Principles of flight simulation*. John Wiley & Sons, 2009.
[14] P. Balakrishna, R. Ganesan, and L. Sherry, "Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of tampa bay departures," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 6, pp. 950–962, 2010.
[15] G. Zhang, M. Mahfouf, M. Abdulkareem, S.-A. Gaffour, Y.-Y. Yang, O. Obajemu, J. Yates, S. A. Soberanis, and C. Pinna, "Hybrid-modelling of compact tension energy in high strength pipeline steel using a gaussian mixture model based error compensation," *Applied Soft Computing*, vol. 48, pp. 1–12, 2016.
[16] C. Fantuzzi and R. Rovatti, "On the approximation capabilities of the homogeneous takagi-sugeno model," in *Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on*, vol. 2. IEEE, 1996, pp. 1067–1072.
[17] J. J. Buckley, "Universal fuzzy controllers," *Automatica*, vol. 28, no. 6, pp. 1245–1248, 1992.
[18] O. Obajemu, M. Mahfouf, and L. A. Torres-Salomao, "A new interval type-2 fuzzy clustering algorithm for interval type-2 fuzzy modelling with application to heat treatment of steel," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 10 658–10 663, 2014.
[19] R. D. Cook and S. Weisberg, "Diagnostics for heteroscedasticity in regression," *Biometrika*, vol. 70, no. 1, pp. 1–10, 1983.