Uimonen, Semen; Tukia, Toni; Ekström, Jussi; Siikonen, Marja-Liisa; Lehtonen, Matti

A machine learning approach to modelling escalator demand response

Published in:
Engineering Applications of Artificial Intelligence

DOI:
10.1016/j.engappai.2020.103521

Published: 01/04/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Uimonen, S., Tukia, T., Ekström, J., Siikonen, M.-L., & Lehtonen, M. (2020). A machine learning approach to modelling escalator demand response. *Engineering Applications of Artificial Intelligence*, *90*, Article 103521. https://doi.org/10.1016/j.engappai.2020.103521

A Machine Learning Approach to Modelling Escalator Demand Response

Semen Uimonen[a,*], Toni Tukia[a], Jussi Ekström[b], Marja-Liisa Siikonen[c], Matti Lehtonen[a]

[a]*Department of Electrical Engineering and Automation, Aalto University School of Electrical Engineering, P.O. Box 15500, 00076 Espoo, Finland*

[b]*Kaiku Health, Vilhonvuorenkatu 11, Helsinki 00500, Finland*

[c]*KONE Corporation, KONE Oyj, Keilasatama 3, P.O. Box 7, Espoo 02150, Finland*

Abstract

This article relates to the topic of the escalator demand response potential. Previous studies mapped escalators as an unrealized potential for additional demand response. The decrease of the nominal speed is the proposed method of reducing the power consumption of an escalator that comes at the cost of passenger travel time and queuing. This work proposes a solution to a problem of selecting appropriate escalators from a large pool to accommodate the target of power curtailment at a minimum cost and highlights the escalator features that constitute the best demand response candidates. The paper compares four methods which differ in calculation speed and accuracy. The primal solution is the earlier developed and enhanced simulation-based model. The random forest and the neural network models provide a solution trained on the output of the simulation-based model aiming to enhance the calculation speed. Finally, all of the developed solutions are compared to the random selection of escalators. The comparison of the proposed statistical approaches shows that the random forest outperforms the neural networks with a maximum error in the prediction of the overall costs in the range of 10.5% of the simulation-based model solution, while the neural network solution lies within 10-58%, depending on the targeted value of the power reduction. Statistical approaches enable performing predictions for different times of the day and for new escalator populations without the need for time-demanding simulations. Comparison to the random selection of escalators demonstrates that the proposed models generally outperform the random selection at least seven-fold.

Keywords: Escalators, Vertical transportation, Demand response, Modelling, Random forest, Neural networks

*I am corresponding author

Email addresses: semen.uimonen@aalto.fi (Semen Uimonen), toni.tukia@aalto.fi (Toni Tukia), jussi.ekstrom@iki.fi (Jussi Ekström), marja-liisa.siikonen@kone.com (Marja-Liisa Siikonen), matti.lehtonen@aalto.fi (Matti Lehtonen)

1. Introduction

There has to be a balance between the generation and the consumption in the power system. The stochastic nature of Renewable Energy Sources (RES) introduces additional challenges for maintaining that equilibrium, while also reducing the system inertia [1]. The ramping numbers of renewable energy generation increase the volatility in the bandwidth of power system frequency.

Besides increasing the energy generation or downregulation by reducing generation, the solutions to mitigate power imbalances include decreasing or increasing the power consumption or shifting the use to off-peak times. This method is called Demand Response (DR). It is used by electric system planners and operators as a resource option for balancing the supply and demand. Usually, DR is directed by financial incentives from the customer perspective. The benefits of DR programs is that they can decrease the cost of electricity in wholesale markets and lead to lower retail prices [2, 3, 4]. Additionally, these can improve the reliability of the power system [3, 5].

While flexible large loads can produce a more meaningful impact on frequency balance, small loads can also participate in the DR events with the help of aggregators. The aggregator is a third party company that contracts with the individual industrial, commercial or residential consumers and aggregates them together. An aggregator company performs as a single DR provider to Transmission System Operator (TSO), Balance Responsible Party (BRP) or to Distribution System Operator (DSO). The individual demand sites can use a combination of increasing on-site generation and/or process shutdown or reduction to deliver the active power demand reduction service. The aggregator receives a percentage of the value created by the reduced power consumption during peak demands, by the balanced intermittent generation or by providing a balancing service [6].

This article has the primary focus on DR using escalator technology. An escalator is the most efficient way to transport large numbers of passengers within a building. There were about 137 000 escalators in the EU in 2016 [7] and about 5000 are installed every year [8]. Escalators can be fixed-speed or intermittent-operating. Fixed-speed escalators are constantly in motion, regardless of the passenger flow. Intermittent-operating escalators are equipped with a Variable-Speed Drive (VSD), which enables energy saving during times when there is no passenger flow. Majority of the newly installed escalators are equipped with the VSD and, since most of the power consumption in the escalator is related to overcoming friction [9, 10], decreasing the speed of the escalator can provide power reduction [11, 12].

1.1. Previous research

In previous research, Al-Sharif described the basis of energy consumption modelling of fixed-speed escalators and explored its dependency on the mechanical systems of the escalators [9]. In [13], Carillo et al. described the benefits of two-speed control in escalators. Intermittent-operating escalators energy consumption was measured and compared to fixed-speed escalators by Kuutti et al. [14]. Articles [10, 15, 16] focused on modelling the energy consumption of intermittent-operating escalators with various passenger volumes and in different traffic patterns and served as a basis for modelling escalator demand response.

Regarding demand response, article [11] describes the potentials of DR for escalator technology in frequency containment markets by means of speed reduction. Frequency containment markets are the reserve markets that serve as a tool for the transmission system operator to obtain additional resources to deal with sudden disturbances or deviations from the balance between power generation and power consumption. It demonstrated that mean reduced power consumption as the result of DR depends on the passenger arrival rates and, thus, the time of the day. Article [12] introduced the costs of escalator DR in terms of increased passengers queuing and travel times.

1.2. Aims and problem statement

Earlier research [11, 12] revealed that escalator DR is a viable option for additional power procurement together with the already established methods. However, DR with escalators provided with speed reduction comes at a cost of slowing the passengers. If the flexibility provided by a bundle of escalators is used in addition to some other established methods, it is not always necessary that all the available escalator flexibility needs to be utilized at all times. Since all the escalator appliances vary by a multitude of parameters, the flexibility and the cost per appliance also vary.

This article aims to provide a robust solution to finding the most suitable options among the existing escalator DR potential at minimal cost. In this study, we first use the previously developed simulation-based model [11] to calculate the flexibility and the cost for each of the available escalators. The downside of the method is the slow calculation speed, which depends on the number of modelled escalators and the amount of necessary calculations or scenarios. Next, we utilized machine learning methods to create statistical models, which drastically increases the speed of computation at the cost of increased prediction error. Furthermore, the article provides recommendations about escalator parameters that are important to prioritize when a larger DR capacity is required, for example when models are unattainable.

In this article, we look at a hypothetical situation where a number of intermittent-operating escalators is available for the DR while the required procurement is a fraction of all the available DR potential. We aim at selecting the escalators for the targeted power procurement so that we maximize the available power reduction of the escalator, while minimizing the increased passenger queuing and travel times. The article compares four solutions: the earlier presented simulation-based model [11], two statistical models and a random selection of available escalators. The two statistical approaches are the neural network and the random forest regression models, which are popular in machine learning. The results are later compared by the calculation speed and the induced error. The structure of the article is the following. Section 2 describes the overall framework of the solution and methodologies applied to it. Section 3 shows the comparison of solutions provided by the research, the statistical models and otherwise a random selection of escalators. Sections 4 discusses the applicability of proposed methods. Section 5 concludes the main results of the article.

2. Methodology

As the case study, we set various targets of necessary power procurement for a simulated group of 4000 escalator units. Since escalators perform differently, we create a pipeline for reaching the solution through modelling and a selection algorithm. The methodological process of arriving to the aforementioned solutions is presented in Fig. 1.

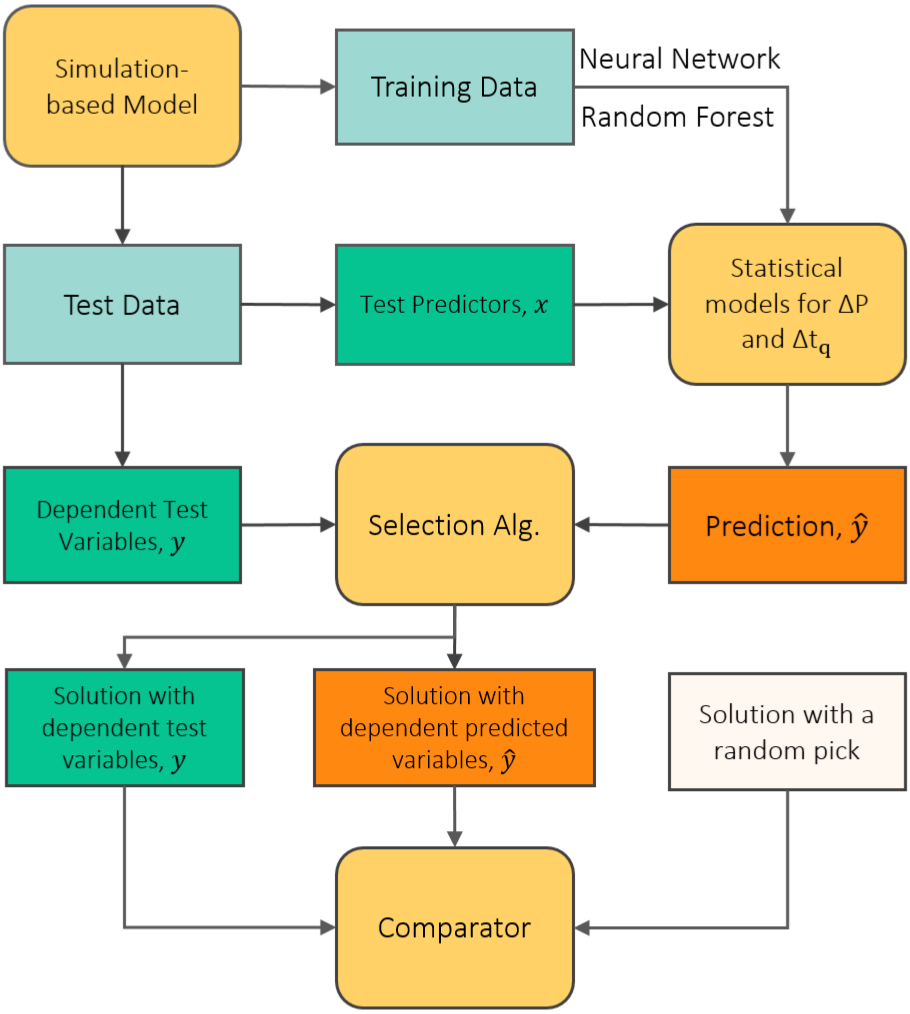

Figure 1: Overall algorithm.

The simulation-based model refers to the previously developed algorithm for modelling the escalators, the aggregate of escalator power consumption and the recovered power during DR events [11]. Subsection 2.1 presents the detailed explanation of the simulation-based model. The model was enhanced with the passenger route choice model, presented later in Subsection 2.2. The outputs of the model are used in the selection algorithm, described in Subsection 2.3. Additionally, in the absence of measurement data, the simulation-based model is the data source for the training and test data necessary to create and verify the statistical models for the predictions of reduced power consumption and increased cost. Statistical models are presented in Subsection 2.4. The detailed data structure is presented in Subsection 2.5.1. Once the predictions and the dependent test variables are obtained, the selection algorithm determines the respective solutions for the data sets. Finally, the obtained solutions from the simulation-based and the statistical models are compared to the random selection solution. The comparison of solutions is presented in Section 3.

2.1. Simulation-based model

The simulation-based model is the data source for the aforementioned statistical model. It is described in detail and is adopted from [11] with added features of loading dependent variable efficiency of the drive and the probability of passengers taking the stairs. Fig. 2 depicts the structure of the model.

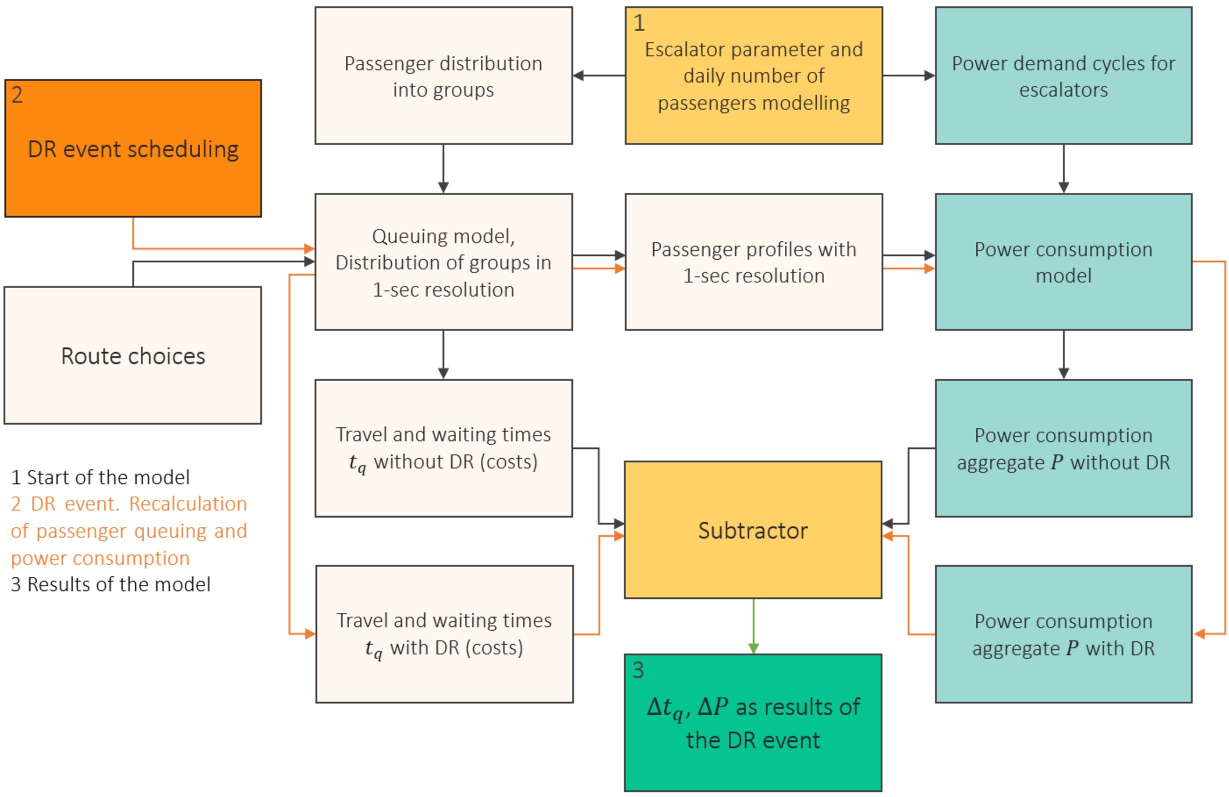

Figure 2: Simulation-based model, adopted from [11].

In this article, the simulation-based model is used to create 12 000 intermittent-operating escalators as training and test data and 4000 escalators as a case study to simulate the DR event lasting five minutes where the escalators are slowed. The VSD technology allows to change the speed of the escalator anywhere from 0 to the maximum allowed. In this article, we have selected 50% of the nominal speed during the demand response event, which is also a typical value during the slow-speed mode of the escalator [10].

Changing the escalator speed raises safety concerns. The highest source of minor injuries on escalators are sudden falls. Oftentimes, falling occurs when the escalator rapidly changes the speed or suddenly stops [17]. Technical standards [18, 19] set limitations in maximum allowed acceleration and deceleration in escalators. In the present article, the authors presume the change of the escalator speed happens within 1 second, which correlates with maximum permitted acceleration, up to 0.5 m/s^2, and deceleration, up to 1 m/s^2, values according to the EN 115-1 2010 [18] standard.

The modelling approach involves creating each escalator with unique parameters that affect the power consumption and passenger travel times. The parameters used in the modelling process are described in Table 1.

Table 1: Modelled escalator parameters, adopted from [11].

Parameter	Description
Type	Fixed-speed or intermittent-operating escalator
Segment	Public transportation or commercial building
Direction	Upwards or downwards
Regeneration	10% of modelled downwards escalators have regeneration
Number of passengers	Daily number of passengers [11]
Energy class	Normally distributed classification indicator that describes the impact of both the efficiency of active parts and the friction of passive escalator components [20]
Staircase	10% of modelled escalators have a nearby staircase
α [°]	Escalator angle 30 or 35 degrees
H [m]	Vertical height of the escalator [11]
Dimensions A, B, C, D	Dimensional reference values derived from [20]
$\mu_{\text{SB/PB}}$	Friction coefficient of step/pallet
$m_{\text{SB/PB}}$ [kg]	Mass of step/pallet
v_{nominal} [m/s]	Nominal speed
t_{travel} [s]	Calculated travel time at nominal speed
η_{nl}	Escalator efficiency at no load (no passengers)
m_{chain} [kg/m]	Mass of chain band per meter

Once the escalators have been created with the daily number of passengers that board the escalator during the modelled hours, passengers are divided into groups according to the building segment passenger distribution. Later, the passengers are redistributed along the timeline with the help of the queuing model to respect the escalator transportation capacity in 1-sec modelling resolution [11]. Current model iteration includes additional calculation of route choices for passengers on escalators that also have a staircase, described in Subsection 2.2. One of the outputs of the passenger queuing model from Fig. 2 is the travel time, t_{travel} and the increased travel and queuing of passengers, denoted as t_q. Escalator travel time at speed v can be expressed with Eq. 1:

$$t_{\text{travel}} = \frac{H}{v \cdot sin\alpha} \tag{1}$$

where H, v, α are the height, the speed and the angle of the escalator from Table 1.

Since the passenger capacity rate of an escalator is proportional to its speed, v, the number of passengers that are in the queue during time t can be expressed with Eq. 2 [11]:

$$N_q(t,v) = \begin{cases} \lambda(t) + N_q(t-1,v) - \mu_{\text{cap}}(v), \ N_q(t,v) > \mu_{\text{cap}}(v) \\ 0, \ N_q(t-1,v) \leq \mu_{\text{cap}}(v) \vee t = 1 \end{cases}, \tag{2}$$

where $t \in [1:86400]$, $v \geq 0$, $\lambda(t)$ is the arriving amount of passengers in time t and $\mu_{\text{cap}}(v)$ is the escalator passenger capacity per second.

The produced 1-sec resolution passenger traffic profiles serve as the input for the power consumption model. The power consumption model produces the aggregate power consumption profiles for all the modelled escalators, denoted as P.

The described modelling outputs are produced for two scenarios, one where there is no DR event taking place and one where there is a DR event of a 5-min length. During the modelled 5-min DR event, every escalator is slowed to 50% of the nominal speed. The speed reduction curtails the power consumption, while on the other hand, it increases the passengers travel and queuing times.

The result of the described model is the difference between the two scenarios, the amount of recovered power per escalator and the associated cost, or the increased travel and queuing time, as the consequence of the DR event, denoted as ΔP and Δt_q. The disadvantage of the described model is the calculation time. It took on average 41 min 17 seconds to calculate a set of results for one simulated day with a specific parameter combination for 4000 escalators, which can be referred as an iteration. Numerous iterations of the simulation-based model with varying time of the DR event produce the training and test data for the developed statistical model in this article. The data is described in Subsection 2.5.1.

2.2. Route choice model for units with a staircase

For escalators that are modelled with a staircase, the route choice calculation for passengers was adopted into the existing model. Article [21] investigated the behaviour of the passengers when they have to choose between using the escalator and the staircase during traffic peak hours. The route choice measurements were conducted in transportation sector and yielded the equations of the route choice probability. It is possible to adopt this approach for modelled escalators in the transportation sector as the measurements in [21] were conducted on escalators with similar average parameters as were used in this article. It is assumed that each pedestrian can evaluate the delay caused by the escalator and choose to remove oneself from the queue with a given probability if walking on the staircase is faster. Probability of a passenger choosing the stairs for upwards- and downwards-moving escalators is calculated according to Eq. 3, 4 [21]:

$$\mathbb{P}_{\text{stairs up}} = \frac{1}{1 + \exp\left(-5.3441 - 0.2073\Delta t_{\text{delay}}\right)}, \tag{3}$$

$$\mathbb{P}_{\text{stairs down}} = \frac{1}{1 + \exp\left(-3.1001 - 0.1745\Delta t_{\text{delay}}\right)}, \tag{4}$$

where Δt_{delay} is the measure of discomfort, which is represented as the difference in time that the passenger would spend standing in the queue and during his journey on the escalator and walking on the stairs.

As a result, passengers behave differently depending on the direction they are heading. They are less sensitive to a relative delay on upwards-running escalators [21].

2.3. Selection algorithm

One of the aims of this article is to produce an algorithm that helps to infer the best fitting escalators for the DR and pick enough escalators to fulfill the targeted power curtailment amount. As mentioned in Section 2.1, the outputs of the simulation-based model after modelling the DR event are the curtailed average power consumption ΔP and the increased passengers travelling and queuing time Δt_q. Finding the best fitting escalators for DR requires selecting them from the pool of all escalators according to a convenient metric. It should aim to maximize the reduced power, while minimizing the cost. For this reason, we have created a metric for escalators, denoted $S = \frac{\Delta P}{\Delta t_q}$ or *score*, which shows the increment in the curtailed power consumption per second of the increased travel and queuing times. The cost is denoted as $C = \frac{1}{S} = \frac{\Delta t_q}{\Delta P}$.

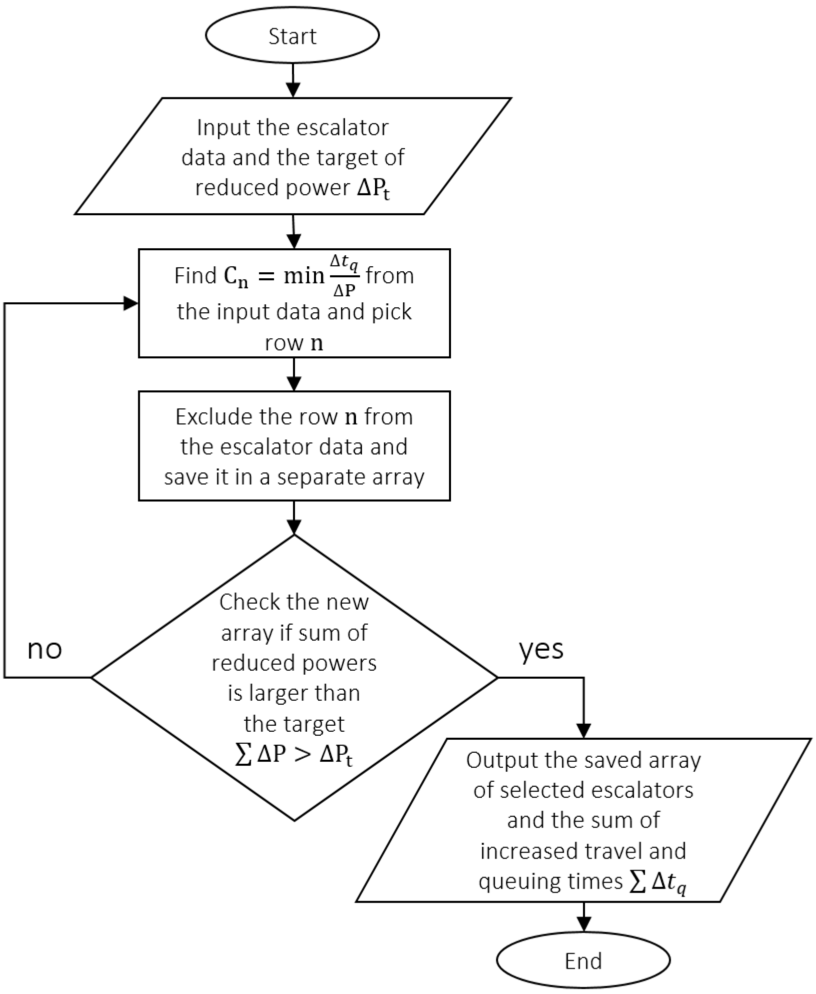

Figure 3: Selection algorithm.

The input data consists of the reduced power consumption per escalators, ΔP, the increased travel and queuing times as a result of slowing down the escalator, Δt_q, the score of the escalator, $S = \frac{\Delta P}{\Delta t_q}$, and the target value for power curtailment ΔP_t. In this article, we chose multiple targets of power curtailment starting with 0.25 MW up to 3.75 MW with steps of 0.25 MW.

At each step, the algorithm finds the minimum value of C at the n-th row from the input data, excludes that row and saves it in the separate array. The process is repeated until the sum of reduced power of the selected escalators reaches the targeted value. The output of the selection algorithm is the set of escalators with reduced power consumption values and the sum of increased passenger travelling and queuing times. Both the simulated data and the predicted values will be used in the selection algorithm for comparison in Section 3.

2.4. Statistical models

There are several advantages of statistical models. There is a drastic increase in the calculation speed and there is no need to model new appliances with the simulation-based model. Additionally, statistical models allow to incorporate the measurement data together with the simulated data to increase the accuracy when more training data is available.

The selection algorithm outputs the most suitable escalators for the proposed DR target. Therefore, the statistical model should predict the input parameters for the selection algorithm, which are the mean power reduction and increased time for each escalator. In this article, we create two regression models, where dependent variables are ΔP and Δt_q. We have selected Random Forest and Artificial Neural Networks as the regression models for both dependent variables. The performance of both approaches is compared in the results section.

2.4.1. Random Forest

Random forests have been successfully applied to various problems in, e.g., genetic epidemiology and microbiology in general within the last five years. Within a very short period of time, random forests have become a major data analysis tool that performs well in comparison with many standard methods [22, 23]. In [24], random forest was used for real-time price forecasting in electricity markets. Article [25, 26] applied the technique to model short-term electrical load forecasting. In [27], random forest was used for solar power forecasting. The technique became popular because it can be applied to a wide range of prediction problems, even if they are nonlinear and involve complex high-order interaction effects. Furthermore, adding to their benefit, random forests produce variable importance measures for each predictor variable [23].

Tree-based methods, in a generalized way, partition the feature space into a set of rectangles, and then fit a simple model, for example a constant, in each of them. Random forest is a modification of *bagging* where a large number of de-correlated trees is created and then averaged. Bagging is a technique of reducing the variance of an estimated prediction model by averaging, in our case, many unbiased trees. Each of them plays the role of a nonlinear mapping from complex input spaces into continuous output spaces. The non-linearity is achieved by dividing up the original problem into smaller ones, solvable with simple models. A split node in the tree maintains a test that is applied to a data sample to send it toward the left or the right child node. The tests are picked by some criteria to group the training samples into clusters where a good prediction can be achieved by simple models [28]. If trees are grown sufficiently deep, the captured interaction structures in the data have relatively low bias [22].

In regression analysis, random forests are formed by growing decision trees depending on a random vector Θ_b, independent and identically distributed (i.i.d.) from the past random vectors $\Theta_1, ..., \Theta_{b-1}$, such that the tree predictor $h(\mathbf{x}, \Theta_b)$ takes on numerical values. The decision tree will traverse down by splitting at each step into a subset of two until it reaches a leaf. The subsets are chosen to minimize either the mean squared or the mean absolute errors. Vector Θ_b characterizes the b-th random forest tree in terms of split variables, cutpoints at each node, and terminal-node values. The output predictions $\hat{y}$ are also numerical and it is assumed that the training set is independently drawn from the distribution of the random vector $Y, \mathbf{X}$ [29].

Random forest grows trees by recursively selecting $m \leq p$ input variables at random as candidates for splitting, where p is the maximum number of input variables. In this article, the value for $m = p$ for both models. The random forest predictor is formed by taking the average over B of the trees $\{h(\mathbf{x}, \Theta_b)\}_1^b$ [29, 22]:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} h(\mathbf{x}, \Theta_k) \tag{5}$$

The mean-squared generalization error of the forest is [29, 22]:

$$E_{\mathbf{X},Y}(Y - E_\Theta h(\mathbf{X}, \Theta))^2 \tag{6}$$

2.5. Artificial Neural Networks

Artificial Neural Network (ANN), or Neural Network (NN), is a nonlinear statistical model, which can be presented as an interconnected group of nodes or neurons. It can be used in a regression or a classification problem. The approach is relatively easy to implement and is popular thanks to its ability to handle non-linear relationships [30]. In article [31], neural networks are used for classification load curves for demand-side management. Article [32] utilizes neural networks for price and energy demand prediction in incentive-based demand response. A neural network is schematically depicted in Fig. 4.

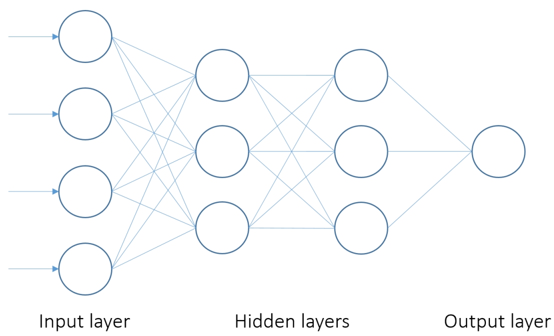

Figure 4: Schematic representation of a NN

A NN can consist of a number of layers labeled as: the input layer, the hidden layer(s) and the output layer. The input layer consists of the features of vector $\mathbf{X}$. The output layer consists of the predicted variable $\hat{y}$. The main element of a NN is the neuron. Each neuron receives its input values from each of the previous layer neurons with the respected weight and bias. The neuron calculates the sum of the waited average of the input values and the bias [33]. For neuron i in layer j, the linear function is calculated according to Eq. 7.

$$z_i^{(j)} = \big(\sum_{d=1}^{k} w_{i,d}^{(j-1)} x_d\big) + b_i \tag{7}$$

where k is the number of neurons in layer $j-1$, $w_{i,d}^{(j-1)}$ is the weight received from neuron d in layer $j-1$, b_i is the bias term. Then, the neuron applies a nonlinear function to $z_i^{(j)}$, which is referred to as the activation function, presented in Eq. 8 as follows:

$$a_i^{(j)} = \varphi(z_i^{(j)}) \tag{8}$$

The activation function is used to determine if the outside connections should consider the neuron to be activated or not, which depends if the calculated value is larger than the activation function threshold [34].

2.5.1. Training, test and case study data

In the article, we use supervised learning to create the statistical models. Data structure is presented in Eq. 9 and Table 2.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \dots & a_{lp} \end{bmatrix} \tag{9}$$

Data sets have a total of $p = 14$ columns, including both independent and dependent variables. The statistical model was created and tested with the data set of $l = 3\ 120\ 000$ rows, where the test data is about 33% of the whole data set, randomly selected. It is a compilation of 12 000 unique escalators where the speed reduction was simulated a total of 20 times for each hour from 8 a.m to 8 p.m.

For comparison of the solutions as depicted in Fig. 1, we use a separate data set with 4000 unique escalators where the speed reduction is simulated once. The case study data set has $l = 4\ 000$ rows. Description of variables columns for all data sets is presented in Table 2. Independent variables are created during the escalator modelling process from the parameters of the simulation-based model, described in Table 1. Dependent variables from the training data set are used in supervised learning during the model training stage to create regression fits.

Table 2: Columns of variables used in the statistical model.

Variable	Variable type
Independent variables	
Segment	Categorical dichotomous
Direction	
Regeneration	
Speed, v [m/s]	
Existence of a staircase	
Angle, α [°]	
Energy class	Categorical ordinal
Daily number of passengers	Continuous
Height, H [m]	
Travel time, t_{trav} [s]	
No load power consumption, P_{nl} [W]	
Hour of DR, t_{DR} [h]	
Dependent variables	
Mean of reduced power consumption, ΔP [W]	Continuous
Increased time, Δt_q [s]	

As mentioned in Fig. 1, the case study data set is split into dependent variables y and independent variables or predictors x. Dependent variables y are fed into the selection algorithm to obtain the solution from the data. It is regarded as the ideal solution to the problem with the existing data and other solutions are later compared to it. Test predictors, created from the data set are used as the input parameters for the statistical model to predict the values of the dependent variables, denoted as $\hat{y}$. Predicted values are also fed into the selection algorithm to obtain the solution with $\hat{y}$. These solutions are compared in Section 3.

3. Results

Article objectives included finding escalator parameters that contribute more to the DR potential of an escalator while minimizing the cost. Additionally, the results section demonstrates the implications of the created models on the overall cost of DR, when the selection algorithm picks the necessary amount of escalators to match the target of power procurement.

3.1. Regression model tuning

Hyperparameter tuning is a vital part of designing a regression model. Presently the values are initially selected empirically [35, 22, 36], further tested and corrected according to the common guidelines for the practitioners, for example such as listed in [37]. Neural networks require to identify the number of layers, the number of neurons in each hidden layer and the activation functions. The selected hyperparameters are presented in Table 3. Table 3.

Table 3: Selected hyperparameter configuration for the neural network model.

Variable	Reduced power	Increased time
Number of hidden layers	1	1
Number of neurons	40	100
Activation function	ReLU	ReLU

Both of the NN models have one hidden layer and rectified linear unit (ReLU) activation functions. The rectified linear unit offers an alternative to the previously most popular sigmoidal [22] nonlinearity function [34]. The rectified linear function is mathematically presented the following way:

$$h_i = max(\mathbf{w}_i^T x, 0) \tag{10}$$

$\mathbf{w}_i^T$ is the weight vector of i^{th} hidden unit and x is the input. In many applications, neural networks with rectified nonlinearities outperform sigmoidal NNs in error metrics and across depth [34, 38].

For random forests it is necessary to choose the number of trees, the maximum number of features and the maximum depth of the tree. The selected parameters are presented in Table 4.

Table 4: Selected hyperparameter configuration for the random forest model.

Variable	Reduced power	Increased time
Number of trees	400	1600
Max. depth	70	None
Max. features	12	12

The default values of the maximum number of features is often set to $p/3$ for regression, where p is the number of predictor variables [22]. In the proposed models, the selected configuration of maximum number of features is set to the value of p. In [39], lower error rates were observed for higher values of maximum selected features for both classification and regression problems. The study recommends to set the value high if there are few relevant variables out of many, so that the algorithm can find those.

3.2. Model fits comparison

To demonstrate the results of the regression models, the correlation between $\hat{y}$ and y can be presented in a scatter plot. The created fits for the mean reduced power and increased times from predictions are presented in Fig. 5 for the 5-min DR event.

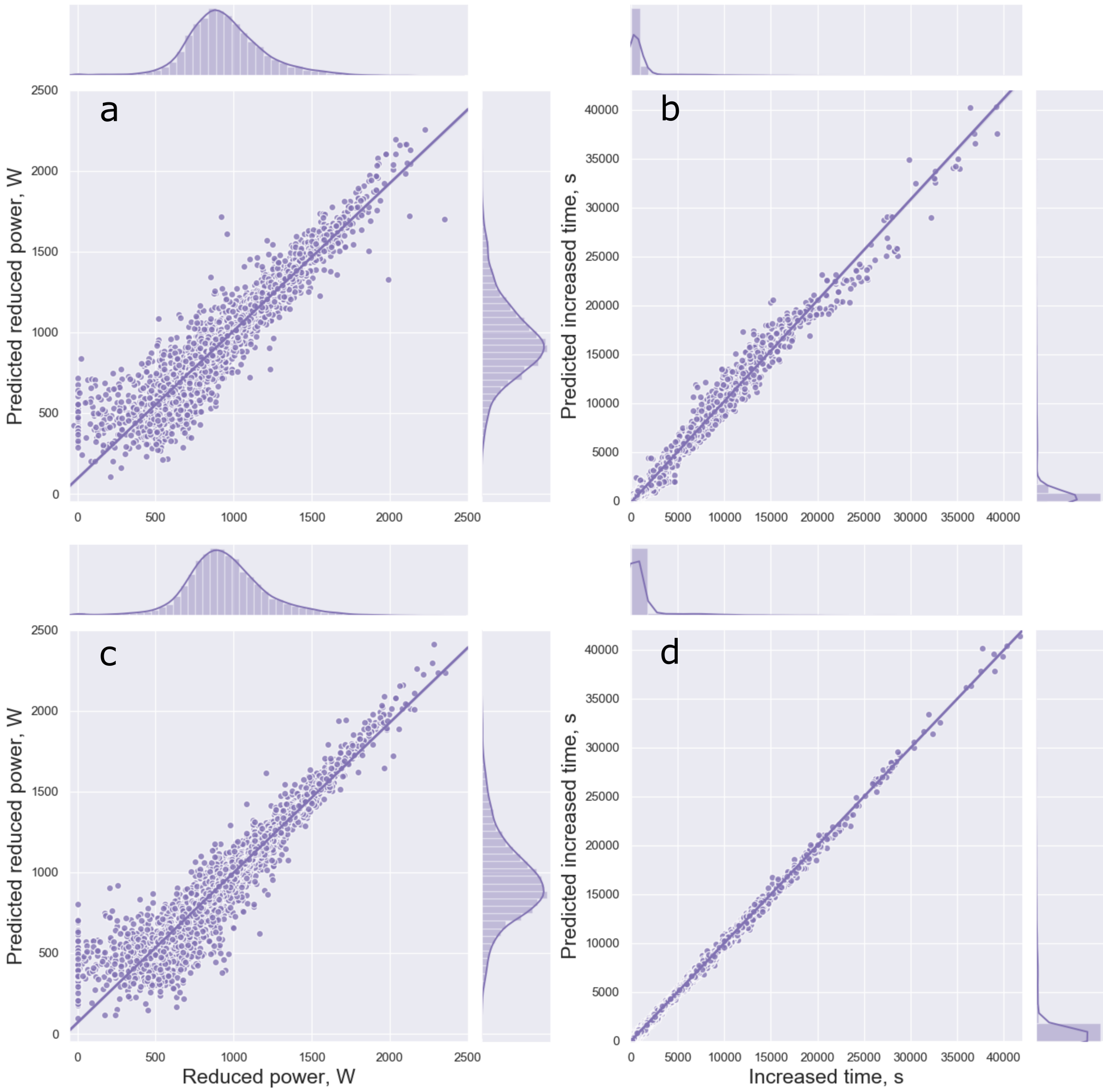

Figure 5: Scatter plots of the dependent variables vs their predicted values with a,b- Neural Network and c,d- Random Forest, sample size 10 000.

We use the R^2 coefficient of determination and the mean-squared errors (MSE) to compare two models, before comparing the solutions in the case study. Table 5 presents the comparison statistics for the models. Both fits have a high Pearson's correlation coefficient and the fits are considered satisfactory. However, the random forest regression model outperforms the neural network in both R^2 and MSE.

Table 5: Comparison of statistical models.

Variable	Increased time		Reduced power	
Method	R^2	MSE	R^2	MSE
Random Forest	0.99	25051.65	0.92	5629.57
Neural Network	0.98	250535.46	0.90	7081.03

The two approaches are less different in predicting the average of the reduced power, while prediction of increased passenger time varies significantly. Thus, random forest is expected to provide more accurate solutions on the case study.

Another approach to compare the results of the two statistical algorithms is to compare the model results for statistical significance. Study [40] shows that the 5x2cv paired t-test has the least probability of producing a type I error. The test performs 5 times the 2-fold cross-validation, where, in each repetition, the available data is randomly split into two equal-sized sets, S_1 and S_2. Each of the learning algorithms, A and B, is trained on each of the sets and is tested on the other, correspondingly. This process yields four error estimates, $p_A^{(1)}$ and $p_B^{(1)}$, trained on S_1 and tested on S_2 and $p_A^{(2)}$ and $p_B^{(2)}$, trained on S_2 and tested on S_1. Estimated differences, $p^{(1)}$ and $p^{(2)}$, are calculated by subtracting the corresponding errors. The estimated variance is then calculated the following way:

$$s^2 = (p^{(1)} - \bar{p})^2 + (p^{(2)} - \bar{p})^2 \tag{11}$$

where $\bar{p} = (p^{(1)} + p^{(2)})/2$. If $s_i^{(2)}$ is the variance computed from the i-th iteration the 5x2cv $\bar{t}$-statistic is calculated the following way:

$$\bar{t} = \frac{p_1^{(1)}}{\sqrt{\frac{1}{5}\sum_{i=1}^{5} s_i^2}} \tag{12}$$

where under the null hypothesis, $\bar{t}$ has approximately a t-distribution with 5 degrees of freedom [40].

The results of the neural network and the random forest regression models were compared with a sample of 500 000 data points for statistical significance with the 5x2cv paired t-test. Two hypotheses are created: H_0 - the null hypothesis and H_1 - the alternative hypothesis. The null hypothesis states that there are no significant differences in the results of two predictors. The alternative hypothesis states that there are significant differences in the two models. The probability is referred to as the p-value. Table 6 shows the calculated t-statistic and p-values. The threshold for accepting the null hypothesis is 0.05.

Table 6: 5x2cv paired t-test statistic and calculated p-values.

Variable	Increased time	Reduced power
t	-6.501	-5.812
p-value	0.001	0.002
Accepted hypothesis	H_1	H_1

The p-value is smaller than 0.05, thus, we reject the null hypothesis and accept that there is a significant difference in the two models in each case.

3.3. Random forest feature importance

The random forest allows to extract feature importance values from the created models. To portray which features contribute the most to the predictions, Fig. 6 depicts feature importance for both reduced power and increased time statistical models.

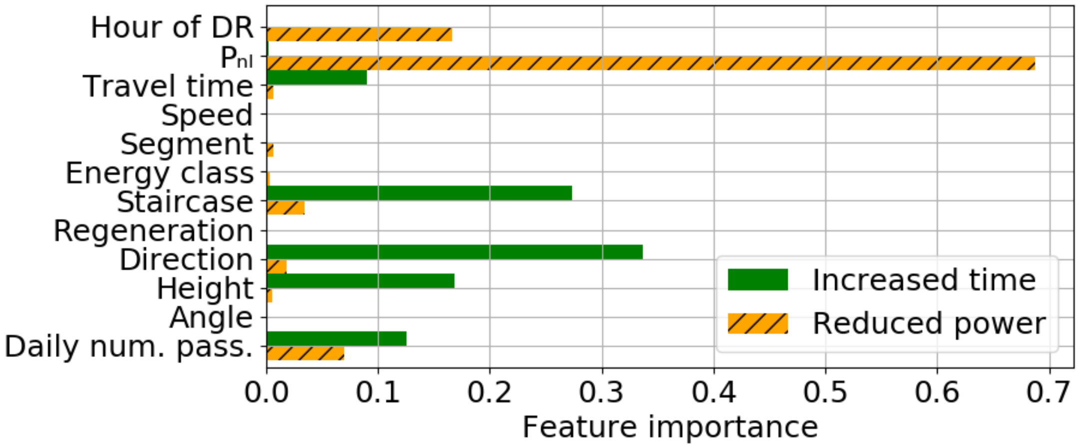

Figure 6: Feature importance for reduced power and increased time statistical model.

Feature importance is derived by ranking the features according to how much they improve the model fit across all created trees on average. However, in situations where there are multiple highly correlated features, some of the randomly selected features will be selected first and, therefore become primary and ranked higher, while the highly correlated features, will later be ranked lower, as they already give less improvement. Therefore, interpretation of feature importance from Random Forests must be done with caution [23]. The most important features for predicting the curtailed power and increased time are the no load power consumption, staircase, direction, travel time, height and daily number of passengers. The no load power consumption is the most important features for predicting the reduced power. At the same time, features such as speed, height and angle are low in significance, while in reality, the no load power consumption is a function of all of them.

3.4. Escalator parameters that affect DR selection

Creating efficient statistical models often requires eliminating less important features from the training data, which reduces the model complexity and increases the computational speed. Studying the feature importance helps to understand the modelling process and the relations between the dependent and the independent variables.

The most important features that contribute to the selection of escalators by the DR score can be studied by looking at the solution provided by the simulation-based model after the selection process. The following subsections test categorical and continuous variables for their significance.

3.4.1. Categorical variables

Fig. 7 presents the comparison of the ratio of the frequency of categorical independent variables in sets of top 10% of escalators selected for DR and the total population of 4000 units.

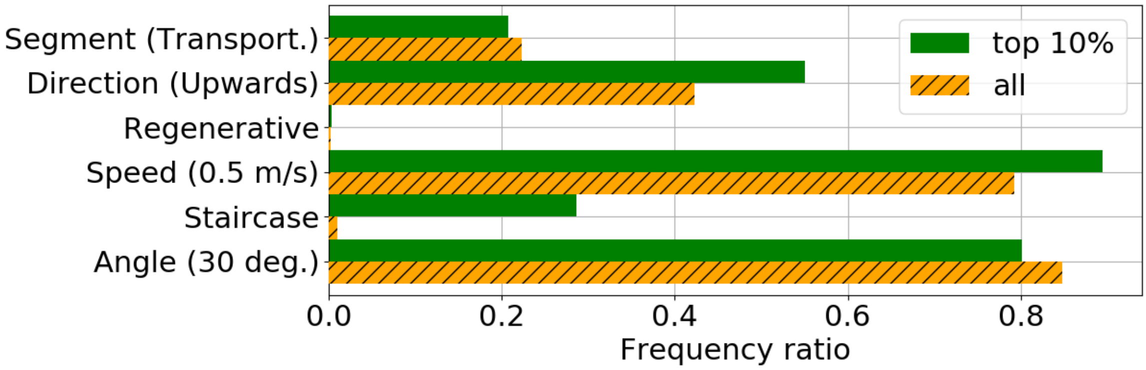

Figure 7: Frequency ratios of categorical variable values in sets of top 10% of escalators selected for DR compared to the total population.

Fig. 7 shows that, from categorical variables, the top 10% of the best escalators for DR has a larger ratio of upwards-running escalators, escalators with the speed of 0.5 m/s and escalators that have a staircase nearby.

We use the Chi-square statistic (χ^2) to determine the likelihood of whether there is a significant difference between the expected distribution and the actual distribution of the categorical variables in the selected samples of data by chance [41]. One set of data is the optimized group of the top 10% of the best escalators for DR, while the other is the overall set of escalators excluding the first set. Samples drawn from the sets should not have large differences in the expected values of frequencies of categorical variables unless these variables have strong ties with dependent variables in the process of the selection of these escalators. Chi-square test aims to find the probability that the distribution of data in the sampled sets is by chance.

The null hypothesis states that the distribution of data in the two sets is different by chance. The alternative hypothesis states that the distribution of data is different between the two sets of data because the tested variables have correlation with the way that the set was formed. In other words it shows that the tested categorical variables affect the selection of best fitting escalators for DR by score, which is correlated with dependent variables of reduced power consumption and increased time.

The calculation of χ^2 statistic is performed with Eq. 13 [42]:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{13}$$

where O_i is the observed value and E_i is the expected value for each category i in categories of $i = 1, 2, 3...j$ of cases, which in our case is $i = 2$, $True$ or $False$.

Table 7 contains the calculated χ^2 statistic and the probability values (p-value) that the null hypothesis is true ($H_0 : O_i = E_i$) for categorical variables in the selected samples of two aforementioned sets.

Table 7: Chi-squared statistic and probability values for categorical variables in sets of the 10% of the best fitting for DR escalators and the rest from the population, where DF is degrees of freedom [41].

Variable	Direction	Segment	Staircase	Regeneration	Speed	Angle
χ^2	29.1399	0.7582	166.074	0.0102	49.3075	5.8643
p-value ($DF = 1$)	6.723e-8	0.3839	0	0.9196	0	0.0154
Accepted hypothesis	H_1	H_0	H_1	H_0	H_1	H_1

Table 7 shows that categorical variables of Direction, Staircase, Speed and Angle have a low probability of H_0 to be true, meaning that these variables are likely to affect the way that the best fitting escalators for DR are chosen by the score. On the contrary, the relatively small difference in the segment and the ratio of regenerative escalators in the selected set of the escalators are by chance with a high probability.

Concluding the categorical variables test and taking into consideration only the modelled parameters, it follows that it is more common for the best fitting escalators for DR by speed reduction to be upwards-running with lower speed (0.5 m/s), have a larger angle and a staircase nearby.

3.4.2. Continuous variables

Table 8 presents the comparison of the mean values, standard deviations and the median values of the continuous variables in the sets of the top 10% of escalators selected for DR and the full set.

Table 8: Comparison of continuous variables mean and standard deviation values in the top 10% of the best escalators for DR and all modelled escalators.

Variable	Mean value		Standard deviation		Median	
	top 10%	all	top 10%	all	top 10%	all
Daily number of passengers	5438.00	7514.00	3705.00	3394.00	3936.00	7096.00
Height [m]	4.15	5.42	1.50	1.79	3.80	5.20
No load power consumption [W]	2166.82	2188.42	484.46	448.68	2048.00	2114.00
Travel time [s]	8.72	11.88	3.58	4.22	8.00	11.00

It is seen that the means of continuous variables except no load power consumption differ in the top 10% of selected for DR escalators and all the data. It is necessary to check if the continuous variables are significant in the selection of the best escalators for DR. Since the variables distributions are non-parametric, we use the Wilcoxon Rank-Sum Test [43].

The Wilcoxon Rank-Sum Test is a non-parametric test which is based on the order in which the observations from the two samples differ. With this test, we wish to test the hypothesis if the distribution and, thus, the medians of the independent variable in two populations are the same. In our case, we have seen that the means and median values of the samples are different.

The null hypothesis states that the distributions of the tested variable in the two samples is the same, thus their medians are the same. The alternative hypothesis states that the distributions are different. Rejecting the null hypothesis would mean that the distributions of the variables are not the same and their medians are shifted. This implies that these variables are significant in how the sample of objects in the group to which we compare the original expected values is formed.

The calculation of the W statistic is performed with Eq. 14 [44]:

$$W = \sum_i [\mathrm{sgn}(x_{2,i} - x_{1,i}) R_i] \tag{14}$$

where the sgn is a sign function that extracts the sign of a real number, $x_{1,i}$ and $x_{2,i}$ are variable values in two groups and R_i is the rank.

Table 9 shows the calculated values for the W statistic and p-values.

Table 9: Wilcoxon sum-rank test statistic and calculated p-values.

Variable	Daily pass. num.	Height [m]	P_{nl} [W]	t_{trav} [s]
W	14.9834	15.2953	1.9467	16.0613
p-value	9.4252e-51	8.2134e-53	0.0515	4.7631e-58
Accepted hypothesis	H_1	H_1	H_0	H_1

The probabilities that the difference in the distributions and the medians of the tested continuous variables, except the no load power consumption, happened by chance is low.

Since the p-value of the no load power consumption in the Wilcoxon sum-rank test was on the border of the acceptance threshold we also test the means of the variables in these two groups by calculating the probability of sampling the same distribution of a variable. If we sample, for example, 10 000 times 400 (10%) escalators from the overall set, the distribution of the mean values of independent variables that we test in these samples will be close to the normal distribution. We can compare the mean values of the selected top 10% of escalators for the DR to the the sampled distribution. We calculate the z-score – the amount of standard deviations (σ) from the mean value (μ) of the distribution of the average values of sampled independent variable and, thus, the probability (p-value) of sampling the selected group of escalators. z-score can be calculated with Eq. 15 [45]:

$$z = \frac{x - \mu}{\sigma} \tag{15}$$

For this test, H_0 - the null hypothesis is that the average of the means of the sampled dependent continuous variables is equal to the mean of the top 10% of the escalators for DR. H_1 - the alternative hypothesis states that the means are different.

Table 10 presents the one sided z-scores and p-values for continuous independent variables for the top 10% of the best for DR escalators being significant for grouping escalators by the score.

Table 10: z-scores and one sided p-values for continuous independent variables for the top 10% of the best for DR escalators being significant for grouping of escalators by the score.

Variable	Daily pass. num.	Height [m]	P_{nl} [W]	t_{trav} [s]
z-score	-12.9757	-14.8572	-1.0227	-15.9067
p-value	8.4036e-39	3.1252e-50	0.1532	2.8482e-50
Accepted hypothesis	H_1	H_1	H_0	H_1

Concluding both tests, the null hypothesis was accepted for the no load power consumption while the rest of the tested variables showed to be significant in the selection of escalators by the DR score presented in this article. Escalators with lower than average daily number of passengers, height and shorter travel time are ranked higher in suitability for DR and are preferable to be selected prior to others.

3.5. Comparison of solutions

To illustrate the comparison of the four escalator selection methods presented in Fig. 1, we compare the increased queuing and travel times for various targets of power curtailment, which was selected in range from 0.25 MW to 3.75 MW with steps of 0.25 MW.

The obtained results from the simulation-based model using the case study data include ΔP and Δt_q. Additionally, we have calculated the score, $S = \frac{\Delta P}{\Delta t_q}$, as one of the input variables for the selection algorithm mentioned in Fig. 1 and 3. For each target of power curtailment, the selection algorithm picks the best fitting escalators according to the score until it fulfills the power consumption curtailment target. In this way, since there is no measurement data of DR events from the field, the results produced by the simulation-based model are considered to be 'ideal' and other solutions are compared to it. The downside of obtaining the results with the simulation-based model is the speed of the calculation. The modelling of one DR event for 4000 escalators takes on average 41 minute 17 seconds. The proposed solution by means of the statistical model trained on the modelling data yields the solution approximately in 0.034 seconds. In comparison to the modelled solutions, randomized selection is the fastest way to choose escalators, however, randomization induces substantial errors. Fig. 8 depicts the comparison of increased travel and queuing times depending on the target of power curtailment for the test sample of 4000 escalators with the compared solutions and the error of the statistical models compared to the simulation-based approach.

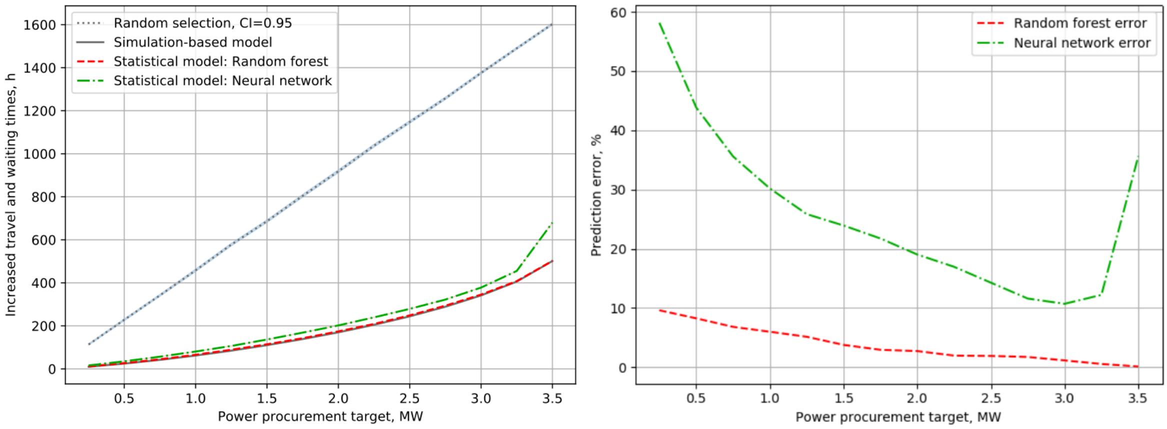

Figure 8: Comparison of methods and statistical model error in increased travel and waiting time compared to the simulation-based model solution.

The random selection line in Fig. 8 is the mean of a 100 random selections for each power procurement target with confidence interval of 95%. Solution yielded by statistical models greatly outperforms the random selection of escalators. Random forest approach outperforms the neural networks on the case study. Fig. 8 shows that depending on the target of power reduction, the neural network error from the simulation-based solution lies within 58 to 10.5%, while the random forest solution yields an error in the range of 10.5 to 0.2%.

It follows that creating a statistical model for escalator DR poses large benefits to possible future scheduling of DR events. Using the statistical model, it is simple to obtain fast predictions about escalator DR performance for new and existing appliances without the need for long simulations. Additionally, statistical models can be improved with more obtained data from real measurements.

4. Discussions

This article presented a framework and for determining the best fitting escalators for possible DR. The proposed solutions are flexible in regard to the available data and perceived targets. Of course, the simulation-based model is seen as more accurate than the statistical methods, because the latter was trained to replicate the simulation output. To truly compare the accuracy of the two methods would only make sense if real-world data were available. The largest obstacle for creating more accurate models is the data. There is little to no data about passenger patterns, escalator segmentation and vertical rise distribution. The proposed methods can be drastically improved with field measured data. The proposed methods were used as a showcase of a probabilistic framework for the calculation of the potential escalator DR. Probably, performance of the statistical approaches is a subject to change if real measurement data is available, which has to be further tested.

Regarding the recommended escalator parameters that yield the highest escalator score for DR and considering the costs, the ideal situation for reducing the power consumption on the escalator is when it moves with nominal speed but has the least number of passengers on board. In cases of maximum number of daily passengers at the times of the DR, there is a high probability to inflict large costs with decreasing the speed of the escalators. Therefore, the daily passenger number should ideally reflect a more significant knowledge about the density of the incoming passenger flow. This knowledge would increase the prediction accuracy and help to reduce the costs of such DR method.

Compared to the previous research, the simulation-based model was updated with the calculation of the probability of passengers taking a staircase if the escalator had a staircase nearby, when queuing occurred. Furthermore, the efficiency of the drive was made variable due to changes in momentary loading, as depicted in [11].

5. Conclusions

This article presented the framework and the model for determining the best fitting escalators out of a large stock to accommodate the power curtailment target. The earlier developed and enhanced simulation-based model together with the selection algorithm is capable of finding the solution with minimal cost. The method was compared to the introduced statistical models based on the random forest and the neural network regression models which produce results faster but with induced errors that are larger for smaller targets of power procurement. The error ranges from 10.5 to 0.2% for random forest and 58 to 10.5% for neural network approach compared to the simulation-based model value. Thus, with the current data and simulation, the random forest is a more accurate and preferable approach for rapid prediction of DR potential and induced cost. In the case of random selection, the cost in terms of increased travel and queuing times is times larger than with the proposed methods. An alternative to modelling of DR is to use the following recommendations derived from the models. While taking into consideration the modelled escalator parameters, the best candidates among the modelled escalators for DR by speed reduction are upwards-running escalators with slower speed (0.5 m/s), with the larger angle and that have a staircase nearby. They also have lower than average daily passenger numbers, height and travel time while having an average no load power consumption.

Bibliography

[1] A. Alahäivälä, Harnessing demand response for power system flexibility, School of Electrical Engineering and Automation, Aalto University, Finland, Ph.D. thesis (2017).

[2] A. Arteconi, D. Patteeuw, K. Bruninx, E. Delarue, W. D'haeseleer, L. Helsen, Active demand response with electric heating systems: Impact of market penetration, Applied Energy 177 (C) (2016) 636–648 (2016). doi:10.1016/j.apenergy.2016.0.

[3] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, V. Miranda, Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois, IEEE Transactions on Sustainable Energy 4 (1) (2013) 250–261 (Jan 2013). doi:10.1109/TSTE.2012.2215631.

[4] M. Aunedi, P. . Kountouriotis, J. E. O. Calderon, D. Angeli, G. Strbac, Economic and environmental benefits of dynamic demand in providing frequency regulation, IEEE Transactions on Smart Grid 4 (4) (2013) 2036–2048 (Dec 2013). doi:10.1109/TSG.2013.2258047.

[5] Y. Zhou, P. Mancarella, J. Mutale, Modelling and assessment of the contribution of demand response and electrical energy storage to adequacy of supply, Sustainable Energy, Grids and Networks 3 (2015) 12 – 23 (2015). doi:https://doi.org/10.1016/j.segan.2015.06.001.

[6] Emissions EUETS.com, Demand Side Response Aggregator (DSR Aggregator) (2018).
URL `https://www.emissions-euets.com/internal-electricity-market-glossary/855-demand-24side-response-aggregator-dsr-aggregator`

[7] Statistical Committee Report, report, Statistical Committee - European Lift Association (2016).
URL `http://www.ela-aisbl.org/images/Statistical_Committee_ELA_GA_2016_02.pdf`

[8] Lifts and Escalators 2016, report, Statistical Committee - European Lift Association (2017).
URL `http://www.ela-aisbl.org/images/Summary_sheet_2016_v2.pptx`

[9] L. Al-Sharif, Modelling of escalator energy consumption, Energy and Buildings 43 (6) (2011) 1382 – 1391 (2011). doi:10.1016/j.enbuild.2011.01.014.

[10] S. Uimonen, T. Tukia, M.-L. Siikonen, D. Lange, C. Donghi, X. L. Cai, H. Nousu, H. Hakala, M. Lehtonen, Energy consumption of escalators in low traffic environment, Energy and Buildings 125 (2016) 287 – 297 (2016). doi:10.1016/j.enbuild.2016.05.018.

[11] S. Uimonen, T. Tukia, M.-L. Siikonen, M. Lehtonen, Potential of aggregated escalator loads in demand response, Electric Power Systems Research 175 (2019) 105917 (2019). doi:https://doi.org/10.1016/j.epsr.2019.105917.

[12] S. Uimonen, T. Tukia, M. Lehtonen, Vertical transportation demand response: Cost of lost customer hours, in: 15th International Conference on the European Energy Market (EEM), 2018, pp. 1–5 (June 2018). doi:10.1109/EEM.2018.8469208.

[13] C. Carrillo, E. Díaz-Dorado, J. Cidrás, M. Silva-Ucha, A methodology for energy analysis of escalators, Energy and Buildings 61 (2013) 21 – 30 (2013). doi:10.1016/j.enbuild.2013.01.038.

[14] J. Kuutti, R. E. Sepponen, P. Saarikko, Escalator power consumption compared to pedestrian counting data, in: 2013 International Conference on Applied Electronics, 2013, pp. 1–4 (Sept 2013).

[15] S. Uimonen, T. Tukia, M. Lehtonen, M. L. Siikonen, Modelling the daily energy consumption of escalators with various passenger volumes, in: IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), 2016, pp. 1–5 (June 2016). doi:10.1109/EEEIC.2016.7555401.

[16] S. Uimonen, T. Tukia, M.-L. Siikonen, M. Lehtonen, Impact of daily passenger traffic on energy consumption of intermittent-operating escalators, Energy and Buildings 140 (2017) 348 – 358 (2017). doi:10.1016/j.enbuild.2017.02.026.

[17] L. Al-Sharif, Escalator human factors: Passenger behaviour, accidents and design, Lift Report 32 (6) (2006) 14 (11 2006).

[18] EN 115-2, Safety of escalators and moving walks – Part 2: Rules for the improvement of safety of existing escalators and moving walks, Standard, CEN-CENELEC, Brussels (2010).

[19] ASME A17.1-2010/CSA B44-10 Handbook: ASME A17.1-2010, Safety Code for Elevators and Escalators, CSA B44-10, Safety Code for Elevators, Standard, The American Society of Mechanical Engineers, New York, NY, USA (2010).

[20] ISO 25745-3:2015, Energy performance of lifts, escalators and moving walks- Part3: Energy calculation and classification of escalators and moving walks., Standard, International Organization for Standardization, Geneva, CH (Oct. 2015).

[21] C. Y. Cheung, W. H. K. Lam, Pedestrian route choices between escalator and stairway in MTR stations, Journal of Transportation Engineering 124 (3) (1998) 277–285 (1998). doi:10.1061/(ASCE)0733-947X(1998)124:3(277).

[22] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001 (2001).

[23] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution.", BMC bioinformatics 8 (2007) 25 (02 2007). doi:10.1186/1471-2105-8-25.

[24] J. Mei, D. He, R. Harley, T. Habetler, G. Qu, A random forest method for real-time price forecasting in new york electricity market, in: 2014 IEEE PES General Meeting — Conference Exposition, 2014,

pp. 1–5 (July 2014). doi:10.1109/PESGM.2014.6939932.

[25] G. Dudek, Short-term load forecasting using random forests, in: D. Filev, J. Jabłkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szynkarczyk, S. Zadrozny (Eds.), Intelligent Systems'2014, Springer International Publishing, Cham, 2015, pp. 821–828 (2015).

[26] A. Lahouar, J. Ben Hadj Slama, Random forests model for one day ahead load forecasting, in: IREC2015 The Sixth International Renewable Energy Congress, 2015, pp. 1–6 (March 2015). doi:10.1109/IREC.2015.7110975.

[27] M. Abuella, B. Chowdhury, Random forest ensemble of support vector regression models for solar power forecasting, in: 2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017, pp. 1–5 (April 2017). doi:10.1109/ISGT.2017.8086027.

[28] H. Phan, M. Maaß, R. Mazur, A. Mertins, Random regression forests for acoustic event detection and classification, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23 (1) (2015) 20–31 (Jan 2015). doi:10.1109/TASLP.2014.2367814.

[29] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32 (Oct 2001). doi:10.1023/A:1010933404324.

[30] A. Marinescu, C. Harris, I. Dusparic, V. Cahill, S. Clarke, A hybrid approach to very small scale electrical demand forecasting, in: ISGT 2014, 2014, pp. 1–5 (Feb 2014). doi:10.1109/ISGT.2014.6816426.

[31] M. Macedo, J. Galo, L. de Almeida, A. de C. Lima, Demand side management using artificial neural networks in a smart grid environment, Renewable and Sustainable Energy Reviews 41 (2015) 128 – 133 (2015). doi:https://doi.org/10.1016/j.rser.2014.08.035.

[32] R. Lu, S. H. Hong, Incentive-based demand response for smart grid with reinforcement learning and deep neural network, Applied Energy 236 (2019) 937 – 949 (2019). doi:https://doi.org/10.1016/j.apenergy.2018.12.061.

[33] A. Ng, K. Katanforoosh, Cs229 lecture notes, deep learning.
URL `http://cs229.stanford.edu/notes/cs229-notes-deep_learning.pdf`

[34] A. L. Maas, A. Y. Hannun, A. Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: in ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013 (2013).

[35] B. Baker, O. Gupta, N. Naik, R. Raskar, Designing neural network architectures using reinforcement learning, CoRR (2016).

[36] P. Probst, M. N. Wright, A.-L. Boulesteix, Hyperparameters and tuning strategies for random forest, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9 (3) (2019) e1301 (2019). doi:10.1002/widm.1301.

[37] J. Heaton, Introduction to Neural Networks for Java, 2Nd Edition, 2nd Edition, Heaton Research, Inc., 2008 (2008).

[38] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Vol. 15 of Proceedings of Machine Learning Research, PMLR, Fort Lauderdale, FL, USA, 2011, pp. 315–323 (11–13 Apr 2011).

[39] B. A. Goldstein, E. C. Polley, F. B. S. Briggs, Random forests for genetic association studies, Statistical applications in genetics and molecular biology 10 (1) (2011). doi:10.2202/1544-6115.1691.

[40] T. G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Comput. 10 (7) (1998) 1895–1923 (Oct. 1998). doi:10.1162/089976698300017197.
URL `http://dx.doi.org/10.1162/089976698300017197`

[41] A. Ugoni, B. Walker, The chi square test: an introduction, COMSIG review, Chiropractors and Osteopaths Musculo-Skeletal Interest Group 4 (1995) 61–4 (12 1995).

[42] A. Agresti, An introduction to categorical data analysis, Wiley, New York, 1996 (1996).

[43] C. J. Wild, Chance encounters: A first course in data analysis and inference (2000).

[44] M. L. M. Richard J. Larsen, An Introduction to Mathematical Statistics and Its Applications (6th Edition), Pearson, 2017 (2017).

[45] E. L. Lehmann, Nonparametrics, Statistical Methods Based on Ranks, Springer-Verlag New York, New York, 2006 (2006).