NEW APPROACH FOR MISSING CONSUMPTION RECONSTRUCTION AND VALIDATION OF ENERGY METER DATA

LUIS MARIA ALONSO[1], AURELIEN WATARE[2], LAURENT-DIDIER ROUX[3],
LIONEL TAUNAY[4], CHRISTOPHE SAUDEMONT[1] & BENOÎT ROBYNS[1]
[1]University of Lille, Centrale Lille Institute of Technology Arts et Metiers, France
[2]Substation Control-Command RTE, France
[3]CNER/DP/TCC RTE, France
[4]Direction Stratégique Réseau SNCF RESEAU, France

ABSTRACT

The opening of the railway use all around Europe carried plenty of technical challenges never seen before in the railway field. An important challenge in energy billing filed, is to precisely do the billing of the consumed energy for each railway company in a reliable way. For that, in ground applications, RTE (the French national transmission system operator, who is in charge of the energy metering in railways by law in France) normally installs two energy meters to ensure the reliability of the measurements. Nevertheless, in agreement with the European law, the installation of only one energy meter is mandatory in the new/refurbished rolling stock since 2014. To ensure the correct billing of the railway companies, the data collected by the energy meters must be validated. In other cases no measure is available due to different technical problems, therefore a system that estimates the energy consumptions of those journeys is necessary. In this paper, a validated dataset is used to reconstruct the amount of energy consumed by trains whose energy meter data is missing or not valid. The validated data will be the start point for different estimators of the consumed energy by the trains. This dataset is composed of the information collected from the energy meters (date information and GPS position of the train) and train run data (transit times on fixed points of the line). Different types of estimators for the energy reconstruction are compared with real measurement data from a regional train circulation for a couple of years. Finally, a discussion of the performance of the different estimators will be shown.
Keywords: energy billing, consumption estimation, neural networks, measurement reliability.

1 INTRODUCTION

The creation of the Single European Railway Directive 2012 [1] started the process of the market opening of the railway use all around Europe. This opening carried plenty of technical challenges (interoperability with different electrification systems, different security systems even different track gauge) not seen for a long time in the railway field. Another important challenge is to precisely do the billing of the consumed energy for each railway company in a reliable way. For that, in ground applications, RTE normally installs two energy meters to ensure the reliability of the measurements. However, the European law only expects the installation of one energy meter (norm EN 50463) in the new/refurbished rolling stock since 2014. To ensure the correct billing of the railway companies the data collected by the energy meters must be validated. For that in Alonso [2] a method based on physical models is shown in the aim of validating the measurement done by the energy meters. This validated data-set (which represents around the 25% of all consumptions) is a great starting point to develop different types of estimators of total consumptions. Four different estimators will be shown and compared in this paper in the aim to find the most performant one.

2 AI TO RECONSTRUCT THE MISSING CONSUMPTIONS

When the energy meters give erroneous data or directly does not communicate, a method has to be developed to estimate the energy consumption. The idea of making a prediction of this

consumption based on the physical models presented in Alonso [2] may look as the best idea in a first time. However, the physical model is highly sensitive to different parameters (as the passenger number or the climatology of the journey) that are unknown when the reconstruction of the consumed energy is done. As can be seen in Tables 1 and 2, those parameters generate big variations in the total consumption. A strong relation between the train mass and the minimal consumption can be observed in Table 1. Also, under DC traction current, the determination of the factor of regenerative braking energy acceptance of the substations (the μ factor) is nearly impossible. As in DC the substations are not reversible in power, the calculation of this μ factor has to take into account other trains or other fixed loads (as stations, maintenance centres…) in the surroundings of the analysed train as well as his positions respect to these loads. Even more, as shown in Andre-Philippe et al. [3] (a study of consumptions based on real data made by SNCF), only the factor of the driving style of the train driver can make to vary the total consumption up to a 30% under DC traction current and near a 12% under AC traction current (whose substations are reversible in power naturally).

Table 1: Impact of the train mass in the minimal and maximal consumption.

Total mass	Minimal consumption	Maximal consumption
Minimal mass, 0 person/m^2 (–24%)	295.88 kWh (–20%)	834.02 kWh (–4.24%)
Standard mass, 4 person/m^2	356.54 kWh	869.44 kWh
Maximal mass, 6 person/m^2 (+10%)	390.70 kWh (+9.5%)	882.71 kWh (+1.5%)

Table 2: Impact of wind in the minimal and maximal consumption.

Wind	Minimal consumption	Maximal consumption
Heavy storm 60 km/h	423.10 kWh (+18.67%)	922.99 kWh (+6.15%)
Light storm 30 km/h	383.88 kWh (+7.67%)	894.51 kWh (+2.88%)
No wind 0 km/h	356.54 kWh	869.44 kWh
Light storm –30 km/h	338.57 kWh (–5.3%)	850.88 kWh (–2.22%)
Heavy storm –60 km/h	329.33 kWh (–8.26%)	838.81 kWh (–3.65%)

The addition of all the impacts in consumption from all the sources of incertitude shows clearly that an estimation based on physical models is far from being the answer to the problem of estimating the consumption in an accurate way. Therefore, other kind of estimators have to be developed; in this paper the performance of four statistical estimators (a mean, a multiple linear regression, a support-vector machine (SVM) and neural networks) will be compared with data from a regional line.

3 CASE STUDY: REGIONAL PASSENGER TRAIN

The studied regional train line is 58.17 km long and has 11 stops. This line has two speed limits, one of 160 km/h for the first 45 km and a second one of 100 km/h for the last 13 km. The line is nearly flat with one end (STA 1) at 7 MASL (metres above sea level) and the other end at 8 MASL (STA 2), the 23rd km being the highest point of the line with 60 MASL. The line is electrified at 1500 V DC with non-reversible substations. In Figs 1 and 2 the repartition of the total consumptions can be observed for years 2016 and 2018, even if the
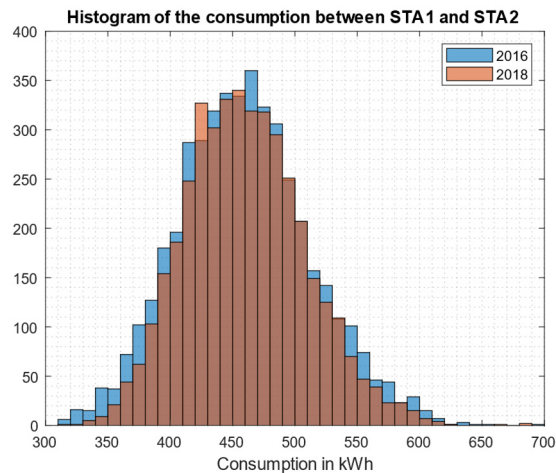
Figure 1:   Consumption histogram for 2016 and 2018 for the journey between STA 1 and STA 2.
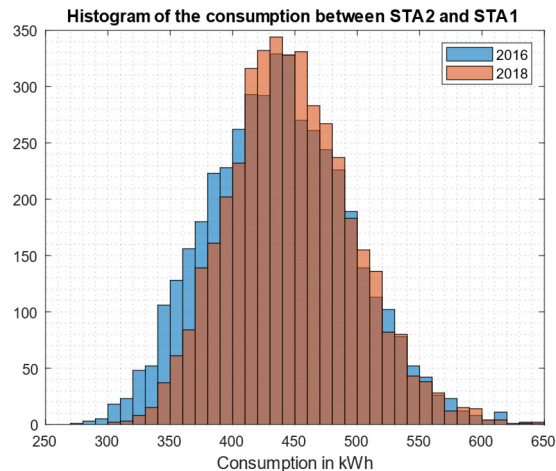

Figure 2:   Consumption histogram for 2016 and 2018 for the journey between STA 2 and STA 1.

majority of the consumptions are between 400 kWh and 500 kWh a gap of a double can be found between the minimal consumption and the maximal consumption.

As mentioned before four different estimators will be compared. For all estimators the same procedure has been followed. A cross validation with a kfold (the number of groups into which data is divided) of 10 has been followed in this procedure. All the data is divided in 10 groups randomly, then nine groups will be used to determinate the estimator values and the performance will be tested over the group that has not been used for the estimator training. This procedure is repeated 10 times changing each time the group where the estimator is

tested. Finally, an average performance is calculated with all the 10 performances already calculated.

For each type of estimator 100 tries are calculated, to evaluate the performance of each indicator the mean error of the 100 tries is calculated as well as the standard deviation of this error.

3.1 Mean estimator

To compare all the estimators a benchmark has to be established. For that the mean has been chosen as it is quite simple to calculate. Two type of mean estimator have been calculated. The first one takes the consumptions of all the year and applying the cross-validation method, this is the "year mean" indicator in Table 3. A second indicator has been developed with the same cross-validation method, but in this case a mean is established for each month of the year. This indicator is the "month mean" indicator in Table 3. In Tables 3–6 the percentage shown is the mean error (over a 100 tries) of the estimator with between parenthesis the standard deviation of this error.

Table 3: Mean estimator performances for the case study.

Estimator	STA 1–STA 2 2016	STA 1–STA 2 2018	STA 2–STA 1 2016	STA 2–STA 1 2018
Year mean	11.57% (0.003%)	10.34% (0.002%)	12.82% (0.002%)	10.97% (0.002%)
Month mean	10.13% (0.009%)	10.06% (0.011%)	11.71% (0.011%)	10.69% (0.013%)

3.2 Multiple linear regression estimator

As the main factors of impact in the consumption are unknown (the number of passengers to calculate the mass of the train, the train driver identification or the meteorological conditions) other sources of data have to be found to estimate the impact of those factors. In the following three estimators, the following listed data (calendar data and meteorological data basically) is used to estimate the total consumption of each circulation:

- Weekend or bank holiday (binary)
- Day of the week (1 to 7)
- Month (1 to 12)
- Day of the year (1 to 365)
- Departure hour (by truncation)
- Journey duration (in minutes)
- Temperature (in °C)

Parametric estimators are widely used in engineering to try to estimate a quantity from different input data with a linear relation. In Ahmad [4] an example is shown where from data related to a car (model, type of gearbox, engine size…), their fuel consumption wants to be reconstructed. The Multiple linear regression is the most basic option of parametric estimators. With input data listed before the following performances (Table 4) have been found.

Table 4: Multiple linear regression estimator performances for the case study.

Estimator	STA 1–STA 2 2016	STA 1–STA 2 2018	STA 2–STA 1 2016	STA 2–STA 1 2018
Multiple linear regression	10.29% (0.098%)	9.9% (0.105%)	12.37% (0.112%)	10.79% (0.108%)

3.3 Support-vector machine estimator

The SVM estimator is the first recommended type of AI to try when a model from data is wanted to be created. The SVM integrates non-linearities than can help to get a better result in comparison to a multiple linear regression. This estimator has been already used in other works [5], [6] to estimate the state of charge (SOC) of lithium batteries that have a high non-linear behaviour. In the case study, the SVM estimator gives the following results.

3.4 Neural network estimator

Neural networks are nowadays the most famous AI tool to build model from data. In the electrical domain they have been used widely for different purposes. In Rad et al. [7] neural networks are used to diagnose fault in electric machines, in Hannan et al. [8] they are used to estimate the SOC of batteries and in Khan et al. [9] they are used to forecast the load of the grid in London. All those studies show how with a good database, neural networks are able to create a model of any system. In the case of energy consumption, a good database can be constructed with all the validated measurements from the energy meters. However, the determination of the topology is not obvious. As the goal of the neural network is to build a model from data, the number of layers and the number of neurons per layer will depend on the complexity of the problem that is being modelled.

In the case study a series of tests have been done to try to find the optimal structure for the neural network. A first approach with only one layer where the number of neurons were varied between two and 40 leads to the following result.

In Fig. 3 the evolution of the mean standard error with respect to the number of neurons used. As it can be observed when the number of neurons is not high enough the mean standard error is high because of the underfitting. Meanwhile when the number of neurons is too high, the mean standard error is also high due to the overfitting in this case. An optimal number of neurons if found between 4 and 10 neurons, this result was also found when the data of 2016 was analysed.

A second test was performed to find if a two layers neural network would improve the performance. For that a series of test were done varying the number of neurons of the first and the second layer between one and 20. In Fig. 4 the results can be observed, the optimal configuration for a two layers neural network will be one neuron in the first layer and between 10 and 15 neurons in the second layer. Nevertheless, the obtained performance is nearly the same as the one found for the one-layer configuration neural network. Due to this reason the one-layer configuration has been retained as the best structure, as it gives the same performance and takes less time to train.

Using a one layer with eight neurons neural network with a Bayesian regularization backpropagation training algorithm, the following results were found in the case study.
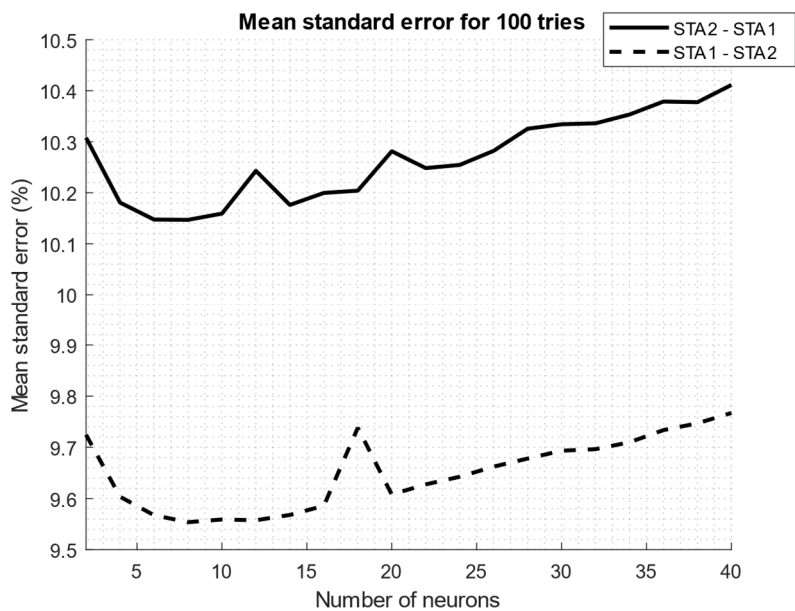
Figure 3: Mean standard error of different configurations of neural networks with data from 2018.
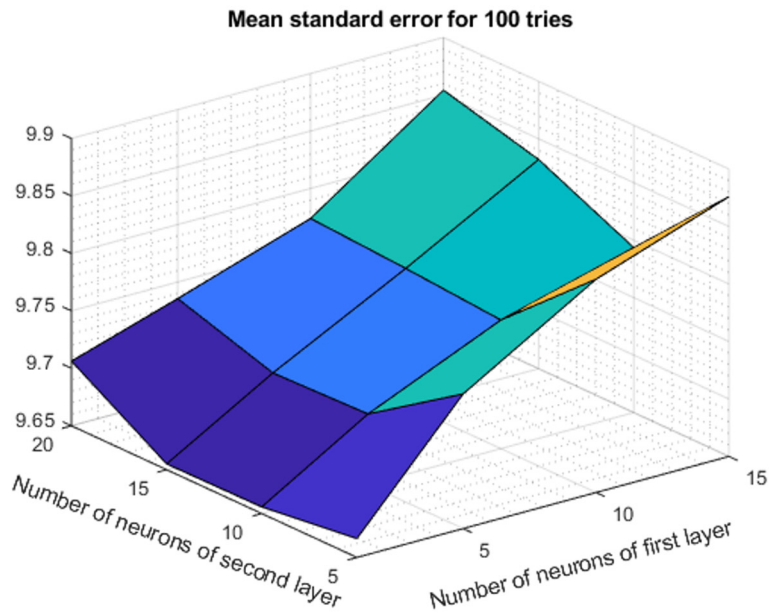
Figure 4: Mean standard error of different configurations of neural networks with data from 2018 for the journey STA1–STA2.

4 RESULT ANALYSIS AND CONCLUSION

If the all results are compared a few conclusions can be reached. First of all, the mean standard error is around a 10–11% which is actually a good precision if all the sources of uncertainty are taken into account. However, when more complex estimators are used (as SVM (Table 5) or neural networks (Table 6)) only a little improvement of around 1–2% is obtained. This means that the estimators are not able to find a correlation between the variations in the total consumption and the input data. This could also mean that the given input data is not valid to estimate the sources of uncertainty (like total mass of the train, behaviour of the train driver…) that have a huge impact in the total consumption

Table 5: SVM estimator performances for the case study.

Estimator	STA 1–STA 2 2016	STA 1–STA 2 2018	STA 2–STA 1 2016	STA 2–STA 1 2018
SVM	9.75% (0.105%)	9.51% (0.132%)	11.43% (0.135%)	10.13% (0.12%)

Table 6: Neural network estimator performances for the case study.

Estimator	STA 1–STA 2 2016	STA 1–STA 2 2018	STA 2–STA 1 2016	STA 2–STA 1 2018
Neural network	9.34% (0.092%)	9.26% (0.098%)	10.71% (0.119%)	9.87% (0.096%)

REFERENCES

[1] EU, Single European Railway Directive, 2012. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0034&from=en. Accessed on: 19 Nov. 2019.

[2] Alonso, L.M., Energy-meter data validation in railway via physical models. *VPPC2020*, 2020. (Submitted.)

[3] Andre-Philippe, C., Didier, F., Pascal, M., Younes, G., Gwenael, P. & Patrick, H., Energy efficiency for rolling stock, from measuring to savings. *2018 IEEE International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC)*, 2019.

[4] Ahmad, N., Specific data model of smart fuel consumption cost estimator. *2009 International Conference on Computer Technology and Development (ICCTD 2009)*, **1**, pp. 429–432, 2009.

[5] Zhang, N. & Liu, K., The predition of SOC based on multiple dimensioned support vector machine. *MACE 2011: Second International Conference on Mechanic Automation and Control Engineering*, **1**, pp. 1786–1788, 2011.

[6] Chen, Y., Long, B. & Lei, X., The battery state of charge estimation based weighted least squares support vector machine. *Asia-Pacific Power & Energy Engineering Conference (APPEEC)*, **1**(2), pp. 1–4, 2011.

[7]   Rad, M.K., Torabizadeh, M. & Noshadi, A., Artificial neural network-based fault diagnostics of an electric motor using vibration monitoring. *Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering, TMEE 2011*, pp. 1512–1516, 2011.

[8]   Hannan, M.A., Lipu, M.S.H., Hussain, A., Saad, M.H. & Ayob, A., Neural network approach for estimating state of charge of lithium-ion battery using backtracking search algorithm, *IEEE Access*, **6**, pp. 10069–10079, 2018.

[9]   Khan, G.M., Khattak, A.R., Zafari, F. & Mahmud, S.A., Electrical load forecasting using fast learning recurrent neural networks. *Proceedings of International Joint Conference on Neural Networks. International Joint Conference on Neural Networks*, pp. 1–6, 2013.