

Energy Consumption Prediction with Uncertainty Quantification for Electric Truck Operations: A Data-Driven Approach

Rik Litjens¹^a, Róbinson Medina²^b, Nikos Avramis²^c, Camiel Beckers²^d,
Steven Wilkins²^e and Mykola Pechenizkiy¹^f

¹*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, the Netherlands*

²*Powertrains Department, TNO, Helmond, the Netherlands*

Keywords: Energy Consumption, Prediction Model, Uncertainty Quantification, Battery Electric Vehicle, Fleet Management, LSTM.


Abstract: The adoption of electric trucks in commercial applications is growing due to the adoption of zero-emission zones in large cities. However, the usage of these trucks shows challenges for fleet managers due to their limited range and uncertain energy usage. Accurately predicting the energy consumption of these vehicles is crucial for their optimal usage in commercial applications. This work introduces a novel energy consumption prediction method for electric trucks, based on a data-driven approach. The approach uses a two-stage Long Short-Term Memory (LSTM) architecture: the first stage predicts vehicle speed while the second predicts energy consumption. For the second stage, two updates to the LSTM encoder are proposed. The first improves the energy prediction by splitting the predictions into regenerated and consumed energy, whereas the second provides a score that quantifies the prediction uncertainty using Student's t -distribution. Evaluating the approach using real-world truck-operation data shows that splitting the energy consumption into regenerative and consumed components contributes to a 20% reduction of error compared to a state-of-the-art LSTM model, mainly due to improved prediction accuracy for regenerated energy. Finally, the t -score demonstrates a 92% reduction of calibration error compared to a Gaussian equivalent. This ensures reliable application in the design of worst-case planning scenarios, decision thresholds, and probabilistic planning approaches.


1 INTRODUCTION


Governments worldwide are committing to reducing greenhouse gas emissions and air pollution to prevent climate change and increase living conditions. A total of 33 countries have pledged to enable 100% zero-emission medium- and heavy-duty vehicle sales by 2040 (DriveToZero, 2024). With a 35% increase in global electric truck sales between 2022 and 2023 and a close to threefold increase within Europe (International Energy Agency (IEA), 2024), it is clear that this is not just a paper reality. In the Netherlands, as in many countries worldwide, municipalities have agreed to introduce zero-emission zones


between 2025 and 2030 as part of the 2019 Climate Agreement (Ministry of Infrastructure and Water Management, 2024; Dutch Government, 2019). The result is that the operation of traditional fossil-fuel-based transport is partially or fully restricted, increasing the importance of emission-free transportation.


For large transportation companies, the introduction of electric medium- to heavy-duty vehicles poses challenges to fleet managers due to the limited range of electric trucks, the difficulties in estimating how much energy is needed to carry out a trip (Pelletier, 2019), and the lack of (fast) charging opportunities along the road. Currently employed solutions introduce inefficiencies in the planning process, by for example oversizing the vehicle battery needed for a trip, or providing conservative energy estimations per trip. An accurate energy consumption prediction would enable the fleet manager to optimize the routes and schedules. However, improving prediction accuracy is often challenging due to uncertainties in the pre-

^a <https://orcid.org/0000-0003-0674-1435>

^b <https://orcid.org/0009-0001-2214-6153>

^c <https://orcid.org/0009-0007-3345-1018>

^d <https://orcid.org/0000-0002-3383-1092>

^e <https://orcid.org/0000-0001-9498-2321>


^f <https://orcid.org/0000-0003-4955-0743>

Table 1: Overview of features and uncertainties for the problem setting.

Available Input Features	Measured Features	Underlying Uncertainties
Gross combined weight (kg)	Front axle speed (km/h)	Wind speed and direction
Coordinates (lat, long)	Brake pedal position (%)	Road conditions
Speed limit (km/h)	Acceleration pedal position (%)	Driver behaviour
Ambient temperature (deg C)	Energy consumption (Wh)	Traffic variability
Elevation (m)		Powertrain efficiency fluctuations
Cornering angle (degrees)		
Traffic lights (binary)		

dictions, given by variations in the route, e.g., due to traffic flow, differences in driver behavior, or weather conditions. Quantifying inherent uncertainties underlying the predictions enables uncertainty-aware decision making in the planning of the vehicles.

Approaches for energy consumption prediction of electric vehicles can be divided into physics-based and data-driven approaches. Physics-based approaches rely on fundamental physical laws to estimate vehicle energy consumption (Yang, 2014; Wu, 2015; Fiori et al., 2016; Fotouhi, 2021). These models have difficulties in capturing variability in energy prediction because of the above-mentioned uncertainties.

Data-driven approaches leverage large datasets and machine learning algorithms to capture the energy dynamics (De Cauwer, 2017; Basso, 2019; Nan, 2022; Pan, 2023). The latter have garnered increasing attention and adoption over the past decade due to their ability to learn relations without the need for the direct specification of vehicle dynamics (Chen, 2021b) and the increasing amount of available data as a whole. This work focuses on data-driven models as they have the potential of capturing the uncertainties in the energy estimations, while still combining insights from the physics-based modeling approaches.

Data-driven state-of-the-art approaches use a two-stage model architecture, where the speed profile and the energy consumption are predicted using Long Short-Term Memory (LSTM) models (Nan, 2022; Chen, 2021a; Petkevicius, 2021; Feng, 2024). However, this technique blindly applies the data-driven approach, where no insights are taken from the application domain, i.e., the physics behind the vehicle dynamics. Some works have touched upon the topic of uncertainty quantification (Petkevicius, 2021; Thorgeirsson, 2021) by evaluating uncertainty scores as a Gaussian random variable. However, the uncertainty score in (Petkevicius, 2021) produces calibration intervals that often do not contain the true values. To the best of the authors' knowledge, no strategy available in the literature leverages the extensive physics knowledge of the electric vehicle

to produce better predictions and quantify the uncertainty in the energy prediction of an electric vehicle reliably.

The contributions of this paper build upon the two-stage predictors from the presented literature. This work improves the second stage (energy prediction) by:

1. providing a novel architecture that takes into account the physics of the vehicle, so that regenerative and consumed energy are modeled separately (LSTM-Decom);
2. introducing the Long Short-Term Memory - Student-t Mean-Variance (LSTM-TMV) method that provides uncertainty scores. The method is based on a Student's t -distribution which outperforms the state-of-the-art Gaussian method.

This paper is organized as follows. Section 2 describes the specific setting for which the models are designed and evaluated. Section 3 describes the prediction architecture and models in it. In Section 4, the approaches relating to uncertainty scores are described. Section 5 presents the results evaluated in a real data set of an electric truck. Section 6 closes the paper with conclusions and future work.

2 PROBLEM STATEMENT

This work focuses on the setting where electric trucks transport goods from a Distribution Centre (DC) to one or more supermarkets. Each truck must drive various routes to ensure all supermarkets remain stocked and return to the DC or a nearby charging station afterward. Each part of the route between a DC and/or a supermarket is considered a "Trip". The challenge is to predict the energy consumption of upcoming trips, using the data features available before departure.

Table 1 describes the available features for the considered setting. Many of these features are accessible before the driving phase and are either determined as part of the operating plan (weight and

route) or retrieved through external information systems (e.g., APIs). The remaining features, which cannot simply be retrieved before a trip, are found in the second column of Table 1. They cannot be used as predictors: instead, they are used for model training purposes or to analyse the results. The third column of Table 1 indicates influencing factors that are infeasible to (precisely) measure or forecast and introduce uncertainty underlying each prediction. Trip data is available for two trucks of the same model over the course of two years (2020 and 2021).

The target of this research is to predict the total amount of battery energy e_i consumed in each trip i , measured in watt-hours (Wh), and quantify the combined magnitude of the underlying uncertainties with an uncertainty score.

3 METHODOLOGY

The method proposed in this paper builds on the start-of-the-art data-driven methods for energy consumption prediction and hence includes a two-stage architecture with the segmentation of the trips (De Cauwer, 2017). To do so, in Section 3.1, trips are divided into smaller segments, then, in Section 3.2, the speed is predicted for these segments and, finally, the energy consumption is estimated.

3.1 Trip Segmentation

The trips are segmented to capture the local driving conditions and road variations that directly impact energy consumption, as well as to provide a consistent input size to the machine learning models. It is assumed that there are no long-term dependencies between data points within the trip, as driving behaviour or events in the past generally only influence the current energy consumption when they occurred recently. This gives the possibility of segmenting the trips into smaller sub-sequences. Furthermore, this structure allows the model to learn from segments on a more fine-grained level, enabling generalization across trips. Note that the prediction targets have been transferred from trip- to segment-level, with each segment j in trip i having its energy consumption prediction target $e_{i,j}$.

Equation 1 formalizes the segmented dataset \mathcal{D} with N trips.

$$\mathcal{D}^{(i)} = ((\mathbf{X}_{i,1}, e_{i,1}), \dots, (\mathbf{X}_{i,j}, e_{i,j}), \dots, (\mathbf{X}_{i,N_i}, e_{i,N_i})) \\ \forall i \in \{1, \dots, N\}. \quad (1)$$

Here, $\mathbf{X}_{i,j}$ represents the sequence of features measured in segment j of trip i , and $e_{i,j}$ is the measured energy. The predicted energy on the segment level is summed as $\hat{e}_i = \sum_{j=1}^{N_i} \hat{e}_{i,j}$, where N_i is the amount of segments in trip i and $\hat{e}_{i,j}$ is the predicted energy for segment j in trip i .

Two methods for segmentation are used in the model pipeline:

1. **Distance Segmenter.** The first segmentation method splits the trips into segments of equal distance S , with a single data point every Δs meters. Δs is defined in such a way that its value represents a forecasting granularity that can be achieved before trips.

This method is used for the speed prediction stage, as time-based methods are unavailable before vehicle speed along the route is known.

2. **Change Segmenter.** The second method is based on the elapsed time within a trip and route characteristics: First, trips are split into segments of T seconds. The choice of T is based on a trade-off between the size of the input segments and the amount of context needed to predict the energy consumption accurately. After this step, the method is extended with additional segment “splits” to make the segmentation occur at more natural points. The technique leverages a subset of input data features that describe the route ahead (traffic lights, maximum speed changes, turns); each time such an “event” occurs, the trip is split into a new segment.

This method is used for the energy consumption stage, as the relation between power and energy consumption depends on time, making the segmentation method more aligned with the physical process.

3.2 Sequential Modeling Architecture

While segmentation narrows the scope of the prediction problem, modular decomposition allows the introduction of more domain knowledge into the design of the model pipeline. As acceleration and speed are influential features in the dataset for predicting energy consumption and they are available as measurements, it is beneficial to create a separate model to predict these features, instead of relating the inputs to the energy consumption directly. The results are subsequently passed to the energy consumption estimation stage. This makes the design of the energy consumption estimation model independent of the design of the speed profile predictor.

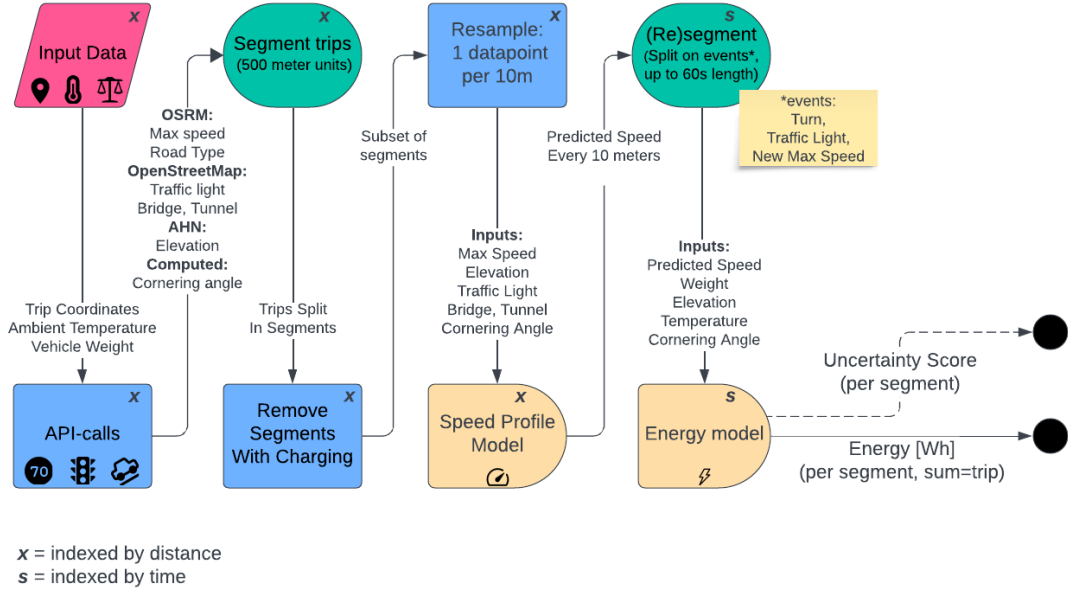


Figure 1: Full evaluation pipeline of the two-stage architecture (uncertainty scores indicated with a dashed line, to indicate they are only present for the uncertainty-aware models). The pink block represents a dataset, the blue boxes indicate (pre-)processing steps, and the green ovals indicate segmentation steps. Online API's include Open Source Routing Mahcine (OSRM), OpenStreetMap, and 'Algemeen Hoogtebestand Nederland' (AHN). Finally, the yellow figures indicate data-driven prediction models and a black dot represents an output.

Formalizing this structure, one can see the segment predictor function f as a composite function consisting of speed profile predictor g and energy consumption estimator h :

$$f(\mathbf{X}_{i,j}) = h(g(\mathbf{X}_{i,j})) \text{ for all trips } i \text{ and segments } j, \quad (2)$$

where function $f: \mathcal{R}^{T_{i,j} \times d} \rightarrow \mathcal{R}$ for d input features and $T_{i,j}$ data points in the input segments.

3.3 Speed Profile Prediction

The first stage of the model aims to predict the speed for upcoming trips. The historical measured speed data can be used as prediction targets during the training process of the speed profile prediction model. Implementing the speed predictor requires a sequence-to-sequence model (i.e., a model that predicts a sequence based on one or more input sequences), such that a single speed prediction is given for each input data point. In this work, speed profiles are predicted for a generic driver, as driver identifiers were not present in the input dataset.

Recurrent Neural Network (RNN) models are a class of sequence-to-sequence models that can learn short- to medium-distance dependencies between the inputs. This aligns with the real-world dependency structure of driving behaviour. Specifically, the model used for this research is a LSTM model, which is an improvement over the original recurrent

neural network that allows learning over longer input sequences. The input features and the target feature are given in Table 2. The initial speed of the segment is provided during the training process and replaced with the last prediction of the previous segment during trip-level predictions.

The model architecture consists of a bidirectional LSTM module, capturing both past and future speed profile indicators. An intuitive example of why this is needed arises when one considers upcoming traffic events or road features, such as traffic lights. The speed profile predictor has to account for these before the vehicle passes the light itself. A stacked 2-layer setup is used to allow more refined processing of the sequence. Each LSTM output is passed through a series of fully connected neural network decoder layers to produce a sequence of predictions.

Table 2: Input and target features for the LSTM speed prediction model. Inputs and predictions are given on a 1Hz basis.

Input Features	Target Feature
Speed limit (km/h)	Speed (km/h)
Elevation (m)	
Traffic lights (binary)	
Cornering angle (degrees)	
Initial speed in segment (km/h)	

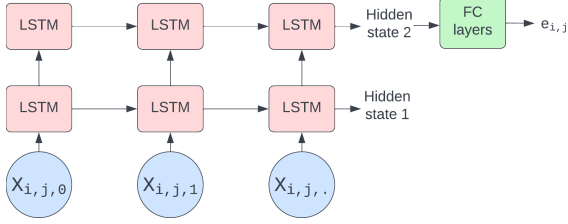


Figure 2: LSTM encoder architecture. The input sequence is encoded before being passed to a series of fully connected (FC) layers that estimate the energy consumption.

Table 3: Input and target features for the LSTM encoder energy consumption model. Input is given on a 1Hz basis, predictions on segment level.

Input Features	Target Feature
Speed (km/h)	Energy (Wh)
Weight (kg)	
Elevation (m)	
Ambient Temperature (deg C)	
Cornering angle (degrees)	

3.4 Energy Consumption Estimation

The subsequent step involves developing a model for estimating energy consumption itself. The training of this model is done independently and involves true speed measurements. Even though the proposed architecture, where second-by-second data within the segments is given as an input sequence, is not found in the literature, the use of LSTM components is widespread in this and other application domains. Therefore, the state-of-the-art model is the LSTM encoder architecture.

The model uses the entire input sequence $\mathbf{X}_{i,j}$ to predict the energy consumption and produces a single scalar output per segment; the input features are described in Table 3. The input is then processed by two stacked LSTM layers, keeping track of two hidden states throughout the process (see Figure 2). At the end of the input sequence, the hidden state of the top layer can be seen as a high-level encoding of the entire input sequence. This encoding is given as input for a series of fully connected layers that produce an energy prediction.

This setup assumes a fixed sequence length for each segment. However, the input segments have varying durations as the Change Segmenter uses events to split the segments. Therefore, the segments are padded to have equal length.

Evaluations of the performance of the LSTM encoder model show that the prediction of the amount of regenerative energy as a result of regen-

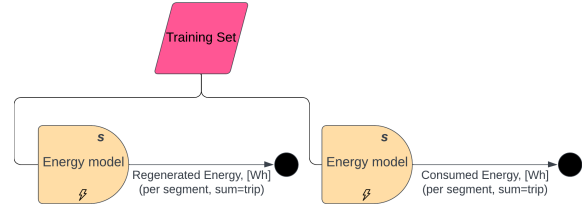


Figure 3: Proposed adaptation on energy consumption pipeline. The entire preprocessing remains the same but two models are trained to predict positive and negative consumption components.

erative braking is the main source of error, as seen in Figure 8. Regenerative braking is a technology in electric vehicles that allows them to recover part of the energy typically lost during braking and redirect it back to the battery. The braking energy recovery is defined within each segment as in Equation 3:

$$e_{\text{regen},i,j} = \int_{t_1}^{t_2} P(t) \cdot \mathbf{1}_{\{P(t)<0\}} dt, \quad (3)$$

where $e_{\text{regen},i,j}$ represents the total recovered energy within segment j in trip i , $P(t) \cdot \mathbf{1}_{\{P(t)<0\}}$ indicates only negative power measurements are summed, and t_1 and t_2 denote the start and end times of the segment. Similarly, the consumed energy $e_{\text{consume},i,j}$ is defined by integrating over all positive power values.

The proposed novel architecture builds on the idea that a different relation between the input features and energy regeneration exists compared to their relation to energy consumption. (Fiori et al., 2016; Chen, 2021a) implemented this idea by training separate models for deceleration and acceleration and swapping between them during inference, as in their setting the actual acceleration state was always known. In practice, exact acceleration states often are unknown. Furthermore, knowing the precise acceleration state is not required in the two-stage segmented model, as predictions are performed on the segment level. Each segment contains both energy regeneration and consumption in varying proportions. The training process of this model is visualized in Figure 3.

A new method called LSTM-Decom is introduced, where two separate models are created that are tasked with predicting the regenerative energy e_{regen} and consumed energy e_{consume} within a segment, respectively. This decomposition allows more control over the design of each model and its training process.

An important design decision for the regenerative model is the use of balanced sampling of the training set. When the input data largely consists of segments with little to no regenerative energy, which is common, the training process is hindered. To implement the balanced sampler, the training set is split into two groups: the first containing segments with more than

a minimum threshold of regenerative energy, and the second group containing the rest. Balanced sampling is used to undersample from the majority group (i.e. the group containing segments with little regenerative energy). For example, consider that there are N_{regen} segments with more than 100 Wh energy regeneration. Then, all N_{regen} high regenerating segments and only $\lfloor N_{regen}/3 \rfloor$ other segments are used as a training set for the energy regeneration model. Besides reducing the size of the training set to allow for faster training, the balanced data set enables a more thorough understanding of energy regeneration.

4 UNCERTAINTY QUANTIFICATION

Even after the proposed model improvement, other sources of error remain, stemming from uncertainties in the input data. To address these errors, a different perspective on model error is useful. Let us define the prediction error with the equation

$$e_{i,j} = h_{\mathcal{D}}(\mathbf{X}_{i,j}) + \epsilon_{i,j} \text{ for all } i, \quad (4)$$

where $h_{\mathcal{D}}$ is the energy consumption model trained on a dataset \mathcal{D} with uncertainties. Viewing $\epsilon_{i,j}$ as a realization of a random variable allows us to account for the inherent uncertainty in the predictions. The aim is to use this new perspective to introduce an additional uncertainty score that describes this error distribution.

An uncertainty score in energy consumption predictions allows fleet managers to have a certain level of confidence in these estimates, helping them anticipate potential variations in energy needs for the vehicles. Decisions that can be made using this score range from the creation of prediction intervals, to worst-case scenarios or the definition of decision thresholds, where the predicted value is only trusted when the accompanying uncertainty is low enough. The uncertainty score is, therefore, expected to follow the properties described in Definition 1.

Remark 1. Expected Properties from an Uncertainty Score Energy Prediction

Two properties are expected from an uncertainty score:

1. A low uncertainty score should correlate with low prediction errors. A fleet manager should be able to place trust in predictions made with high certainty.
2. The uncertainty score should be low when possible. When the model indicates high levels of uncertainty for all predictions, it is impossible to act on it.

Another application of the uncertainty scores entails a probabilistic approach. Consider, for example, a planning application based on probabilities of a truck being able to complete a segment or trip with a certain amount of energy E . This involves calculating $P(e_{i,j} < E)$. To do so, a point-prediction is made using input $\mathbf{X}_{i,j}$, after which a deficiency or surplus is found with respect to E , denoted by $E = h(\mathbf{X}_{i,j}) + \Delta_{E,i,j}$. For the uncertainty-aware models, the theoretical cumulative distribution function (CDF) can be used to derive this desired probability (see Section 4.2 for a concrete example).

4.1 Gaussian Distribution Method

The state-of-the-art methods assume a zero-mean Gaussian distribution for the model error, where the variance is different for each $\epsilon_{i,j}$, i.e.,

$$\epsilon_{i,j} \sim \mathcal{N}(0, \text{Var}_h(\mathbf{X}_{i,j})) \quad (5)$$

with $\text{Var}_h(\mathbf{X}_{i,j})$ as the predicted error variance for input segment j in trip i . The upgrades needed to the LSTM encoder architecture are shown in Figure 4. As both the distribution mean and variance are predicted, the model is called the Gaussian Mean-variance Predictor (GMP).

4.2 Proposed Student's t -distribution Method

This work introduces a novel approach leveraging Student's t -distribution with ν degrees of freedom (t_{ν}),

$$\epsilon_{i,j} \sim t_{\nu}(0, \text{Var}_h(\mathbf{X}_{i,j})). \quad (6)$$

The t -distribution is a probability distribution used in statistics to estimate population parameters when the sample size is small or the population standard deviation is unknown. It features heavier tails compared to the normal distribution, reflecting the increased variability typically present in smaller sample sizes. To facilitate the learning process within the LSTM framework, let us derive from the likelihood function \mathcal{L}_{ν} of the Student's t -distribution with mean 0 and ν degrees of freedom, the Student's t Negative Log-Likelihood (TNLL) as

$$\begin{aligned} \text{Loss}(\epsilon_{i,j}, \text{Var}_h(\mathbf{X}_{i,j})) &= -\log \mathcal{L}_{\nu}(\text{Var}_h(\mathbf{X}_{i,j}) | \epsilon_{i,j}) \\ &= \frac{1}{2} \log(\text{Var}_h(\mathbf{X}_{i,j})) \\ &\quad + \frac{\nu+1}{2} \log \left(1 + \frac{\epsilon_{i,j}^2}{\nu \text{Var}_h(\mathbf{X}_{i,j})} \right) \\ &\quad + \text{constant}, \end{aligned} \quad (7)$$

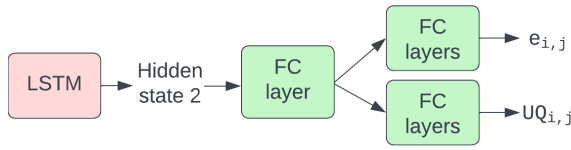


Figure 4: LSTM encoder architecture with uncertainty quantification. The top branch predicts energy consumption, while the bottom branch produces an uncertainty score.

where \mathcal{L} indicates the loss function. As the loss function is purely used for comparison, the constant is irrelevant and removed in the implementation of the loss function. The cumulative distribution function for the Student's t-distribution, used for the probabilistic use case, is:

$$P(e_{i,j} < E) = P(\epsilon_{i,j} < \Delta_{E,i,j}) \\ = I_z\left(\frac{v}{2}, \frac{1}{2}\right), \quad \text{with } z = \frac{v}{v + \frac{\Delta_{E,i,j}^2}{\text{Var}(\mathbf{X}_{i,j})}}. \quad (8)$$

where $e_{i,j}$ is the true energy consumption and E an arbitrary reference value for energy consumption. $I_z(a, b)$ is the regularized incomplete beta function.

The neural network design is equal to that of the Gaussian mean-variance setup, as seen in Figure 4. The core difference is the custom PyTorch loss function incorporating the TNLL loss function. The resulting t mean-variance predictor is named LSTM-TMV.

4.3 Performance Metrics

The desired properties for the uncertainty score, as listed earlier in Definition 1, are translated to two concrete metrics:

1. **Calibration:** This metric captures how often a predicted value is inside the bounds of a prediction interval. For uncertainty-aware models, $(1 - \alpha)$ prediction intervals are constructed, where α represents the significance level. This statistical quantity reflects the fault tolerance by allowing up to α probability of the true value falling outside the prediction interval. The intervals can either be one-sided (to aid in creating a worst-case scenario upper bound or lower bound) or two-sided, in which case both a lower bound and upper bound are present. The percentage of predicted trips whose true value falls within this predicted interval indicates the *calibration*. Theoretically, this rate should equal $1 - \alpha$. Let us therefore define scores where the rate is $1 - \alpha$ or higher as well-calibrated. This metric captures that a “low

uncertainty corresponds to low error”, at least for $(1 - \alpha)\%$ of the predictions. When the uncertainty is lower, the interval is smaller, leaving less room for error. For example: if the prediction intervals for 95 out of 100 test trips contain the true value and α was chosen to be ≥ 0.05 , the model is well-calibrated.

2. **Sharpness:** This metric captures how far the predicted value is from the actual value. In principle, assigning a high uncertainty score to each prediction leads to a high calibration value. This behavior is unwanted. Therefore, we score the adaptivity/tightness of the uncertainty score by generating prediction intervals. The distance between the true energy consumption value for a trip and the bounds on the prediction interval defines the value for this metric. To give an example: if a trip has an energy consumption of 50 kWh and the upper bound is predicted as 55 kWh, the sharpness is 10%.

Note that the uncertainty scores are predicted on segment level, whereas the predictions should be available on trip level. Simply adding the predicted variances produces faulty trip-level scores due to the presence of covariances between individual segments in a trip. To face this issue, a correlation coefficient ρ is introduced. The error covariances between segment j and k in trip i are determined using:

$$\text{Cov}_{j,k} = \rho \sqrt{\text{Var}_h(\mathbf{X}_{i,j}) \text{Var}_h(\mathbf{X}_{i,k})} \quad (9)$$

5 RESULTS AND ANALYSES

This section presents how the models described in Section 3.4 and 4 are trained on a vehicle dataset. To do so, Section 5.1 presents a motivational case study, Section 5.2 shows the results for the novel energy consumption model design, and Section 5.3 for the uncertainty score results. The speed profile predictor results are important to understand the results of the full prediction pipeline. Hence, Section 5.4 shows these results before Section 5.5 describes the full prediction pipeline results.

5.1 Use Case Description and Model Training

A dataset is available containing data collected over two years on two trucks: *Truck #1* and *Truck #2*. Two test sets are used to evaluate the performance. The first set consists of data collected in year 2 for Truck #1 (1123 trips), to show that the methods can learn

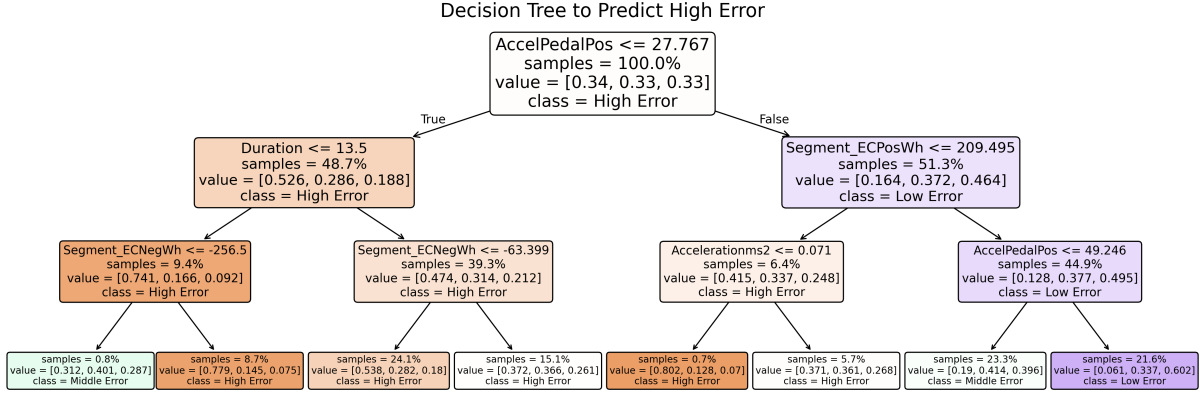


Figure 5: Decision tree to find subgroups of trips with high relative (percentage) error for the LSTM model.

from the past to predict the future. The trips for this truck are, on average, 51 kilometers long, with outliers up until 125 kilometers. The second set consists of data from year 1, for Truck #2 (2748 trips), to show model performance for data measured on a different truck. Note that this truck is of the same type and produced at the same time. The trips for Truck #2 are only 16 kilometers long on average, significantly shorter than for the first truck. For both trucks, trips are recorded in a temperature range of -4°C to 32°C . Elevation difference is only minimally present for all trips.

For the segmentation methods, the chosen parameters are $S = 500\text{ m}$ and $\Delta s = 10\text{ m}$ for the Distance Segmenter, based on the granularity of measured speeds in the dataset. The Change Segmenter used in the energy consumption stage has a maximum duration of $T = 60\text{ sec}$, producing segments that can generalize to other trips while ensuring sufficient context for predictions.

To guide the training process for both stages, the training data set is split into an input set and a validation set. This is done by putting every 8th week of data in the validation set and leaving the rest in the input set. The model that performs best on the validation set (in terms of segment-level mean squared error) is saved and after all 100 epochs, this model is returned as the training result.

The speed LSTM model was trained using a learning rate of 0.0001, a batch size of 64, and a hidden dimension of size 512. For the energy consumption model, the learning rate was 0.0001, the batch size was 64, and the hidden dimension 1024. The threshold for the balanced sampler, used to train the decomposed model was found to be 100 Wh. For the uncertainty models, correlation coefficient ρ was determined to be 0.15, such that the Gaussian two-sided prediction intervals with $\alpha = 0.05$ are well-calibrated. The same parameter is used for other intervals and the

LSTM-TMV model. The best fitting degrees of freedom parameter ν for the t -distribution is 4.

5.2 Energy Consumption Estimator

The results of the standard LSTM encoder model are compared against the decomposed LSTM-Decom model in Table 4 on two test sets. Both models are trained using measured speed data. In terms of all included metrics, the decomposed model outperforms the standard model. A 19% reduction of mean absolute prediction error is observed on the first test set and a 22% reduction on the second. To confirm the motivation behind the decomposed architecture, let us observe the sources of error for the LSTM model by fitting an interpretable decision tree model with depth three and 3 classes: high relative error (assigned to the 33.3% of segments with the highest percentage error), low relative error (assigned to the 33.3% of segments with the lowest percentage error), and medium relative error (assigned to the rest of the segments). The classification accuracy is only 51%, but the method serves as a tool to find high or low-error subgroups in the input dataset. Here, error as a result of noise is assumed to be distributed equally across feature values. The aim is to find subgroups in the input dataset for which the error is systematically higher or lower.

Figure 5 visualizes the resulting tree. The average acceleration pedal position strongly influences whether the error will be high or low, with a low average acceleration pedal position within a segment correlating with higher errors. More concretely, 45% of the segments have a low average acceleration pedal position, and of those segments, 54% have high relative error. For segments with high acceleration pedal positions, the amount of high error segments is only 16%. The reason is most likely that from a certain acceleration pedal position and below, regenerative braking occurs. Figure 6 confirms this by showing

Table 4: Performance metrics comparison for LSTM and LSTM-Decom models across two truck datasets.

Metric	LSTM (default)		LSTM-Decom (proposed model)	
	Truck #1 year 2	Truck #2 year 1	Truck #1 year 2	Truck #2 year 1
MAE (Wh)	1229	1397	978	1083
RMSE (Wh)	2122	2340	1722	1880
MAPE (%)	7.4%	8.5%	6.0%	6.6%
Error <10% (%)	74%	66%	82%	79%

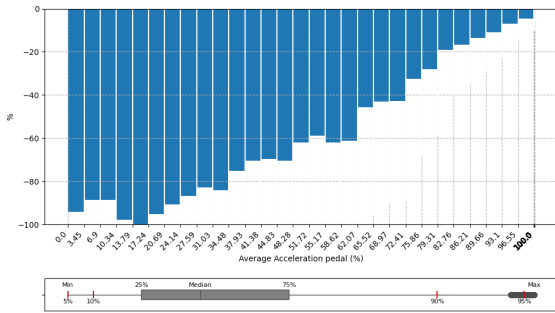
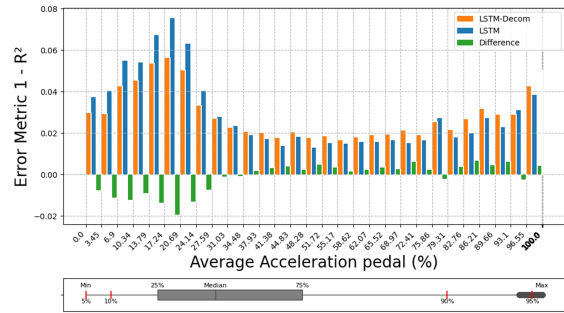


Figure 6: Mean regenerative energy across different values of the average acceleration pedal within segments. The y-axis is normalized between the minimum and maximum values for confidentiality reasons. The boxplot beneath the x-axis indicates the distribution of values within the dataset.

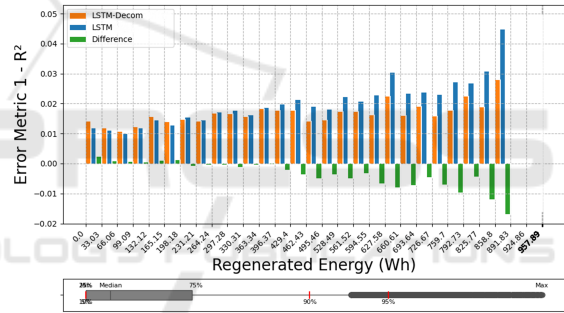
a strong correlation between segments with a low acceleration pedal position and higher amounts of regenerative energy.

The LSTM-Decom model improves the performance specifically for those segments with a low average acceleration pedal position, as shown in Figure 7a. The chosen error metric is one minus the R^2 scores across segments with different average pedal position values. The R^2 is a measure of how much of the variance in the energy consumptions is explained by the inputs. As most values are relatively close to 1 the difference to one is used, implying that low values of this metric indicate good performance.

The improvement in predicting regenerative energy is confirmed by Figure 7b, where increasingly higher regenerative energy correlates with higher improvements in prediction accuracy for the LSTM-Decom model. Even with these improvements, the biggest source of error (derived using the decision tree method) remains the low average acceleration pedal position. This might be related to the fact that the amount of energy that is regenerative is highly uncertain. Overall, these results show that the LSTM-Decom model improves prediction accuracy by leveraging domain knowledge to face the main error source, leading to a 20% reduction of mean absolute percentage error over both test sets.



(a) $1-R^2$ metric across different values of the average acceleration pedal within segments.



(b) $1-R^2$ metric across different values of the average regenerative braking within segments.

Figure 7: Energy consumption estimation comparison.

5.3 Uncertainty Scores

The results of the uncertainty-aware models are first evaluated for the prediction interval application by creating such intervals for confidence terms $\alpha = 0.05$ and $\alpha = 0.01$, indicating low and very low tolerance for errors, respectively. The results in Table 5 provide the results for both one-sided upper bounds and two-sided intervals. The former can be used for the design of worst-case scenarios, whereas the latter gives a more complete image of the quality of the uncertainty score.

Results show that for 6 out of 8 evaluated intervals, the LSTM-TMV model produces well-calibrated scores (see Section 4 for context on the metrics). For the Gaussian model, this only holds for 3 out of 8 intervals, with several calibration values

far from the target (e.g. 75% for the two-sided interval with $\alpha = 0.05$). When looking at the average calibration error (distance from $1 - \alpha$), the error for the GMP model is 7.7%, whereas the average calibration error is only 0.6% for the LSTM-TMV model. This corresponds to a 92% decrease in calibration error. As the calibration metric defines the reliability of the intervals, the LSTM-TMV model is deemed best for modeling the uncertainties. However, this comes at the cost of a higher median sharpness as listed in Table 5. This means that the intervals for LSTM-TMV are generally wider than for the GMP model. When leveraging the score in planning applications, this results in more “conservative” bounds than the Gaussian scores.

Let us look at another use-case for the uncertainty score: a decision threshold, below which predictions are deemed reliable or above which they are marked as too risky. Figure 8 visualizes the relation between model error and predicted uncertainty scores. It shows how a lower bound can be drawn, indicating that lower uncertainty scores correspond to lower model errors. Thus, the figure can aid the end-user in choosing a suitable threshold. Figure 9 shows that the median sharpness of generated prediction intervals using the LSTM-TMV model is higher for lower average acceleration pedal positions in segments. This is consistent with what is expected, considering that many uncertainties are paired with the amount of regenerative in these cases.

To give an example of the probabilistic use-case of the uncertainty score, let us evaluate an example trip selected from the first test set. With a predicted consumption of 20,407 Wh and a predicted standard deviation of 1069 Wh, a fleet manager might wonder what the probability is that 22,000 Wh of energy is enough to complete the trip. Using the CDF function provided in Section 4.2, a probability of 89.5% is derived.

5.4 Speed Profile Predictor

The results of the LSTM model for speed predictions are given in Table 6. For the speed prediction and the full prediction pipeline experiments (Section 5.5) only the test set for Truck #1 in year 2 is used for evaluation.

Figure 10 shows the resulting predicted speed profiles for two sampled trips. Generally, the predicted speed sequences follow the measured speeds, showing the ability to predict acceleration and deceleration patterns. However, the sample on the left is an example of a phenomenon observed throughout the dataset; the model seems unable to capture high-frequency

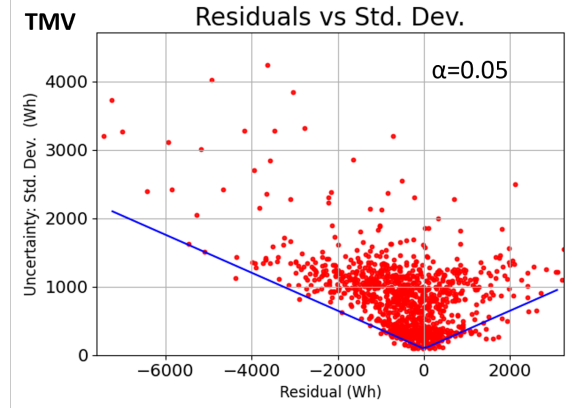


Figure 8: Lower bound on the uncertainty score (st.dev) for all residuals of the LSTM-TMV model on test set “Truck #1 - year 2”. $\alpha = 0.05$ relates to the design of the lower bound, such that 95% of the points lie above it.

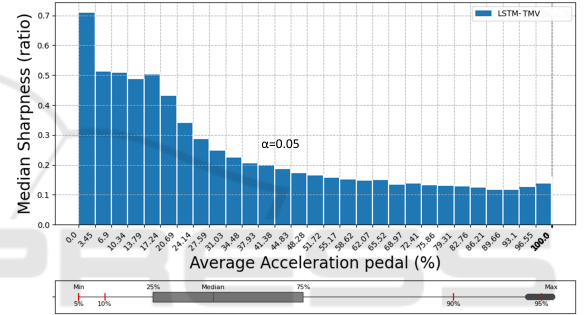


Figure 9: Visualization of the median sharpness (two-sided interval with $\alpha = 0.05$) across different values for the average acceleration pedal position. The uncertainty scores are higher for lower acceleration pedal positions.

changes (seen in the middle of the segment) and rather predicted a smoothed velocity profile. The predicted speeds around 3000 meters of the sample on the right show that certain driving behavior was expected (two peaks) but not measured in the dataset. This shows how the model can learn patterns, which might depend on uncertainties in driving behavior or the environment.

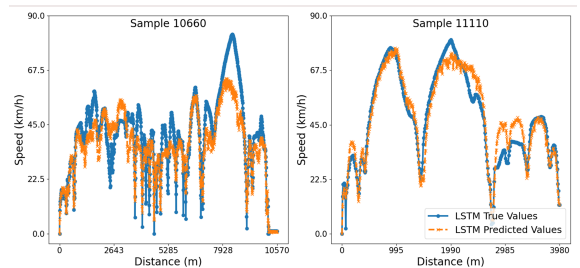


Figure 10: Predicted speed sequences plotted against measured speed data for two sampled trips in the year 2 for Truck #1.

Table 5: Quantitative analysis comparing the uncertainty-aware LSTM-GMP and LSTM-TMV models. Calibrations meeting alpha thresholds are in bold.

Model Comparison	LSTM-GMP		LSTM-TMV	
Dataset	Truck #1 Year 2	Truck #2 Year 1	Truck #1 Year 2	Truck #2 Year 1
$\alpha = 0.05$:				
Calibration upper bound	96.44%	97.82%	95.90%	93.67%
Calibration two-sided	75.07%	76.20%	95.37%	95.92%
Median Sharpness	10.22%	10.92%	11.14%	11.86%
Median Sharpness two-sided	24.27%	25.93%	29.01%	30.90%
$\alpha = 0.01$:				
Calibration upper bound	98.13%	99.20%	98.84%	99.53%
Calibration two-sided	91.54%	89.56%	99.11%	99.75%
Median Sharpness	14.40%	15.39%	19.58%	20.85%
Median Sharpness two-sided	31.82%	34.00%	48.11%	51.23%

Table 6: Speed profile results.

Metric	Truck #1 Year 2
MAE (km/h)	3.64
RMSE (km/h)	5.59
MdAPE (%)	7%

Table 7: Results on the full prediction pipeline for the LSTM-Decom and LSTM-TMV models (Trip-level).

Metric	LSTM-Decom	LSTM-TMV
MAE (Wh)	1507	2253
RMSE (Wh)	2484	3658
MAPE (%)	10.8%	12.50%
Error <10% (%)	59%	40%

5.5 Full Prediction Pipeline

Evaluating the speed profile predictor in combination with both the LSTM-Decom model as well as the LSTM-TMV model leads to the results in Table 7. The LSTM-Decom model shows a mean absolute prediction error of 10.8% and predicts 59% of segments with an error of 10% or below, showing suitable performance for a majority of segments. Compared to the evaluations for the energy consumption models in Section 5.2, higher errors are observed for the full prediction pipeline. Given the presence of prediction errors for the speed model, this decline in performance is expected. The calibration of the LSTM-TMV model in the full prediction pipeline as seen in Table 8 is slightly below the target rate for the predicted intervals and alpha values. This is most likely because the uncertainty in the predicted speed is not accounted for. Still, the calibrations are generally high, indicating that the scores convey meaningful information about the underlying uncertainties.

Table 8: Calibration and sharpness results for Truck #1 year 2 at different alpha values.

Metric	Truck #1 Year 2
$\alpha = 0.05$:	
Calibration upper bound	90.45%
Calibration two-sided	77.61%
Median Sharpness	11.63%
Median Sharpness two-sided	30.29%
$\alpha = 0.01$:	
Calibration upper bound	95.72%
Calibration two-sided	94.56%
Median Sharpness	20.44%
Median Sharpness two-sided	50.23%

6 CONCLUSIONS AND FUTURE WORK

Motivated by the increasing adoption of battery-electric trucks and their operational challenges related to energy consumption prediction, this work investigated a two-stage data-driven approach to predict the energy consumption of electric trucks. The approach integrated a speed profile predictor and an energy consumption estimator.

The key findings of this paper improve the energy consumption estimator or provide a measure of the uncertainty in the predictions. In the first key finding, the energy consumption estimation is improved by including the novel Long Short-Term Memory (LSTM)-Decom architecture, which decomposes the energy consumption into regenerative and consumed energy. The second key finding produces an uncertainty measure in the energy prediction. Compared to the state-of-the-art Gaussian uncertainty scores,

the implemented LSTM-TMV approach assumes a t -distributed error, which successfully produced well-calibrated prediction intervals.

The approach is evaluated using a real data set recorded in electric trucks. Results show a mean absolute prediction error of 7.4%, when evaluating only the energy estimation part (i.e., using true speeds). The reduction of error compared to a standard LSTM encoder architecture is 20%, where analysis shows this is due to improvements in independently predicting regenerative energy. Evaluating the uncertainty quantification scores, the novel t -distributed error approach reduces the calibration error (when compared to a Gaussian approach) by as much as 92%.

The resulting approach shows a mean absolute prediction error of 10.8%, when both the speed and energy consumption are estimated (i.e., the combined pipeline). The decrease in the prediction error compared to state-of-the-art techniques and the provided uncertainty in prediction error make the approach suitable for planning operations.

Future work approaches focus on improving the pipelined by including probabilistic approaches to predict speed, exploring uncertainty propagation from speed to energy prediction, and enhance the data-driven predictions by leveraging from the insights provided by the well-know vehicle physic behaviour.

ACKNOWLEDGEMENTS

This work has received financial support from the Dutch Ministry of Economic Affairs and Climate, under the grant ‘R&D Mobility Sectors’, projects Green Transport Delta - Electrification (GTD-e) and Charging Energy Hubs (CEH), and the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101192657, under the title of FlexMCS.

REFERENCES

- Basso, R. (2019). Energy consumption estimation integrated into the Electric Vehicle Routing Problem. *Transportation Research Part D: Transport and Environment*, 69:141–167.
- Chen, Y. (2021a). Data-driven estimation of energy consumption for electric bus under real-world driving conditions. *Sustainable Transport, Energy, Environment, & Policy*, 98:102969.
- Chen, Y. (2021b). A Review and Outlook of Energy Consumption Estimation Models for Electric Vehicles. *SAE International Journal of Sustainable Transportation, Energy, Environment, & Policy*.
- De Cauwer, C. (2017). A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. *Energies*, 10(5):608.
- DriveToZero (2024). Memorandum of understanding (mou) on zero-emission medium- and heavy-duty vehicles. online.
- Dutch Government (2019). National climate agreement. Policy report on adaptation and mitigation strategies to combat climate change.
- Feng, Z. (2024). Energy consumption prediction strategy for electric vehicle based on LSTM-transformer framework. *Energy*, page 131780.
- Fiori, C., Ahn, K., and Rakha, H. A. (2016). Power-based electric vehicle energy consumption model: Model development and validation. *Applied Energy*, 168:257–268.
- Fotouhi, A. (2021). Electric vehicle energy consumption estimation for a fleet management system. *International Journal of Sustainable Transportation*, 15(1):40–54.
- International Energy Agency (IEA) (2024). Global EV outlook 2024. <https://www.iea.org/reports/global-ev-outlook-2024>. Licence: CC BY 4.0.
- Ministry of Infrastructure and Water Management (2024). Zero-emissiezones in nederland. Web page. Information about the implementation dates and locations of zero-emission zones in 29 Dutch municipalities. Accessed: 01-07-2024.
- Nan, S. (2022). From driving behavior to energy consumption: A novel method to predict the energy consumption of electric bus. *Energy*, 261:125188.
- Pan, Y. (2023). Development of an energy consumption prediction model for battery electric vehicles in real-world driving: A combined approach of short-trip segment division and deep learning. *Journal of Cleaner Production*, 400:136742.
- Pelletier, S. (2019). The electric vehicle routing problem with energy consumption uncertainty. *Transp. Research Part B*, 126:225–255.
- Petkevicius, L. (2021). Probabilistic Deep Learning for Electric-Vehicle Energy-Use Prediction. In *17th International Symposium on Spatial and Temporal Databases*, pages 85–95, virtual USA. ACM.
- Thorgeirsson, A. (2021). Probabilistic Prediction of Energy Demand and Driving Range for Electric Vehicles With Federated Learning. *IEEE Open Journal of Vehicular Technology*, 2:151–161.
- Wu, X. (2015). Electric vehicles’ energy consumption measurement and estimation. *Transportation Research Part D: Transport and Environment*, 34:52–67.
- Yang, S. (2014). Electric vehicle’s electricity consumption on a road with different slope. *Statistical Mechanics and its Applications*, 402:41–48.