

Incorporating Navigation Context into Inland Vessel Trajectory Prediction: A Gaussian Mixture Model and Transformer Approach

Kathrin Donandt

University of Duisburg-Essen

Federal Waterways Engineering and Research Institute

Karlsruhe, Germany

kathrin.donandt@baw.de

Dirk Söffker

University of Duisburg-Essen

Duisburg, Germany

soeffker@uni-due.de

Abstract—Using data sources beyond the Automatic Identification System (AIS) to represent the context in which a vessel is navigating and consequently improve situation awareness is still rare in machine learning approaches to vessel trajectory prediction (VTP). In inland shipping, where vessel movement is constrained within fairways, supplementary navigational context information is indispensable. In this contribution targeting inland VTP, Gaussian Mixture Models are applied, on a fused dataset of AIS and discharge measurements, to generate multi-modal distribution curves, capturing typical lateral vessel positioning in the fairway and dislocation speeds along the waterway. By subsequently sampling the probability density curves of the GMMs, feature vectors are derived which are used, together with spatio-temporal vessel features and fairway geometries, as input to a VTP transformer model. The incorporation of these distribution features of both the current and forthcoming navigation context improves prediction accuracy. The superiority of the model over a previously proposed navigation context-sensitive transformer model for inland VTP is shown. The novelty lies in the provision of preprocessed, statistics-based features representing the conditioned spatial context, rather than relying on the model to extract relevant features for the VTP task from contextual data. Oversimplification of the complexity of inland navigation patterns by assuming a single typical route or selecting specific clusters prior to model application is avoided by giving the model access to the entire distribution information. The methodology’s generalizability is demonstrated through the usage of data of three distinct river sections. It can be integrated into an interaction-aware prediction framework, where insights into the positioning of the actual vessel behavior in the overall distribution at the current location and discharge can further enhance trajectory prediction accuracy.

Index Terms—vessel trajectory prediction, transformer model, multi-source information fusion, Gaussian Mixture Models

I. INTRODUCTION

Predicting vessel trajectories is a pivotal aspect in ensuring the safety and efficiency of autonomous shipping operations.

This publication originates from a joint research project between the German Federal Waterways Engineering and Research Institute (BAW) and the Institute for Ship Technology, Ocean Engineering and Transport Systems (ISMT) (University of Duisburg-Essen), and was developed and written in collaboration with the Chair of Dynamics and Control (University of Duisburg-Essen).

For an autonomous vessel to navigate without collision and effectively, it must possess a comprehensive understanding of the likely future paths of nearby vessels, enabling it to determine the most advantageous course. Anticipating potential collisions allows for timely avoidance maneuvers, while strategic trajectory planning optimizes energy consumption. In the context of inland shipping, characterized by confined navigation channels, fluctuating flow velocities, and shallow water conditions, proactive planning is imperative. This requires a precise assessment of the current navigation scenario and its possible developments. The trajectory prediction is essential for evaluating the options. Inspired by the success seen in other fields, particularly in the realm of automotive trajectory prediction, numerous approaches to vessel trajectory prediction (VTP) in the maritime domain have recently embraced deep learning (DL) techniques, as evidenced by works like [1]–[3]. The efficacy of these prediction models depends largely on the quality and comprehensiveness of the training data. Many DL-based studies on VTP rely solely on AIS data, forecasting future trajectories based on observed spatio-temporal displacement features such as latitude, longitude, speed over ground (SOG), and course over ground (COG). These approaches prove inadequate for inland navigation due to the unknown spatial constraints. Consequently, this study proposes the inclusion of supplementary features that encapsulate spatial conditions which are derived from statistical pre-evaluations of AIS data annotated with hydrological measurements and basic river geometries. The proposed model is provided with “statistical” maps illustrating the discharge-dependent and local distribution of vessel locations in relation to the fairway and of vessel dislocation speeds. By employing Gaussian Mixture Models (GMM) to model these distributions, their inherently multi-modal nature is effectively considered. Aside from solely focusing on individual VTP in this study, disregarding the influence of surrounding ships, the integration of features representing navigational behavior distributions is anticipated to offer significant advantages, particularly in the context of joint VTP. The DL-based VTP research is advancing to approaches considering interacting vessels [4]–[6], which is especially

crucial for the inland navigation domain. By incorporating knowledge of how neighboring vessels behave in relation to the expected behavior derived from distribution-representing features, an enhancement in trajectory prediction performance is expected.

The paper is organized as follows: In Section II an overview of DL-based VTP methods is provided, emphasizing the use of multi-source datasets, the transformer architecture, and the integration of GMMs, contextualizing this contribution within this landscape. In Section III, the proposed methodology, including the generation of supplementary features using GMMs and the transformer-based VTP model, is detailed. In Section IV, the data sources and their fusion are introduced. The evaluation of the proposed model, including an ablation study and comparison with baselines and a previously proposed transformer-based VTP model, is presented in Section V. Finally, the paper concludes with closing remarks, highlighting limitations and suggesting future directions.

II. RELATED WORK

A. Multi-source datasets in DL-based VTP

Few previous DL methods for VTP include supplementary data in addition to AIS. Dijt & Mettes [7] propose a hybrid DL architecture for VTP of inland vessels. Apart from AIS data, the authors use data from radar images and electronic navigational charts (ENCs). A CNN sub-module learns, as a side task, to semantically segment radar images by mapping them to ENCs, and the output of the convolutional layer is used in addition to the SOG and COG values as input to an RNN. Mehri et al. [8] use contextual data in the training data setup step. For the LSTM model, trained to correct the prediction of a physical model for motion prediction, compressed trajectories from AIS with reduced redundancies are used as data source. For a logical consistent compression, shoreline data is considered. Shiptype-specific trained LSTM models are then applied whenever a prediction of the physical model is illogical, i.e. is crossing land or untraversable corridors. Huang et al. [9] fuse AIS and discretized meteorological data including wind, wave, ocean current, and ocean temperature information, and train a VTP model combining convolutional neural networks with multi-head attention mechanism. Finally, Donandt et al. [10] use fairway boundary data and waterway kilometer (KM) information to define the dislocation of inland vessels in relation to a river-specific coordinate system, and train a transformer encoder-decoder to predict future trajectories of inland vessels.

B. Transformer models for VTP

In line with prior developments ([5], [7], [11]–[16]), VTP is modeled as a sequence-to-sequence task, employing an encoder-decoder DL architecture. Departing from the prevalent LSTM- or GRU-based encoder-decoder models, a transformer encoder-decoder model [17], gaining attention in the VTP research recently [1], is chosen instead due to its parallel computation and success in modeling complex sequence-to-sequence tasks. One of the first works applying transformers

to VTP is the TrAISformer [18], first proposed in [13], where discretized latitude, longitude, SOG, and COG features are used in a classification framework. The future trajectory is predicted by iteratively feeding back the last prediction to the model, increasing the input sequence with each prediction time step. More recently, Huang et al. [9] use solely the transformer encoder part (self-attention) to learn the dependency between the features obtained from convolutional modules processing the fused AIS and meteorological data. The future trajectory is, however, generated by a further CNN module, with the transformer encoder serving only as a sub-module rather than the main VTP model. Donandt et al. [19] introduce a context-sensitive classification transformer encoder-decoder, CSCT, which re-frames the regression as a classification task as proposed in [13]. The authors use a vector describing the upcoming river context in terms of curvature for the initialization of the decoder to enable spatial situation awareness. Unlike [13], the entire output sequence is generated at once by the transformer decoder based on the encoded observation sequence, only considering previous predictions in the decoder itself through masking, as originally proposed in [17]. Finally, the recent work of Liu et al. [20] introduces a spatio-temporal multi-graph transformer network to simultaneously predict the trajectories of interacting vessels. Transformer components are used to encode the individual vessel trajectories and the graph representation of the traffic scene, while the prediction of the vessel trajectories are accomplished with an LSTM.

C. Gaussian Mixture Models in the context of DL-based VTP

Gaussian Mixture Models are applied in Murray & Perera [21] to cluster historical trajectories. Based on the cluster of trajectories to which a selected vessel's trajectory is assigned, a trajectory prediction is obtained via a dual linear autoencoder model. In contrast to the usage of GMM in a prior step to the VTP task, Dalsnes et al. [22], fit a GMM on the multiple possible trajectories obtained from the proposed Neighbor Course Distribution Method (NCDM). Thus, a probabilistic model of the future positions is obtained, accounting for uncertainty and multi-modality.

This study presents an extension of previous approaches by integrating hydrological data, i.e. discharge, with AIS data to enhance the spatial situation awareness, particularly for the inland shipping domain. Different from earlier approaches delegating the extraction of pertinent features for VTP from contextual data sources such as maps to the model, the prior fusion of contextual data with vessel driving patterns is proposed. Unlike previous works using transformer components, a holistic approach is adopted by employing the entire transformer encoder-decoder model for the prediction of future time steps. Instead of the formerly used classification reframing in transformer-based approaches, the (dis-)location features are not discretized any more, resulting in a reduction of model parameters. The output consists of parameters of a bivariate Gaussian distribution, thus allowing approximation of the uncertainty [11]. Moreover, GMMs are employed to generate a kind of cluster information in the form of sampled

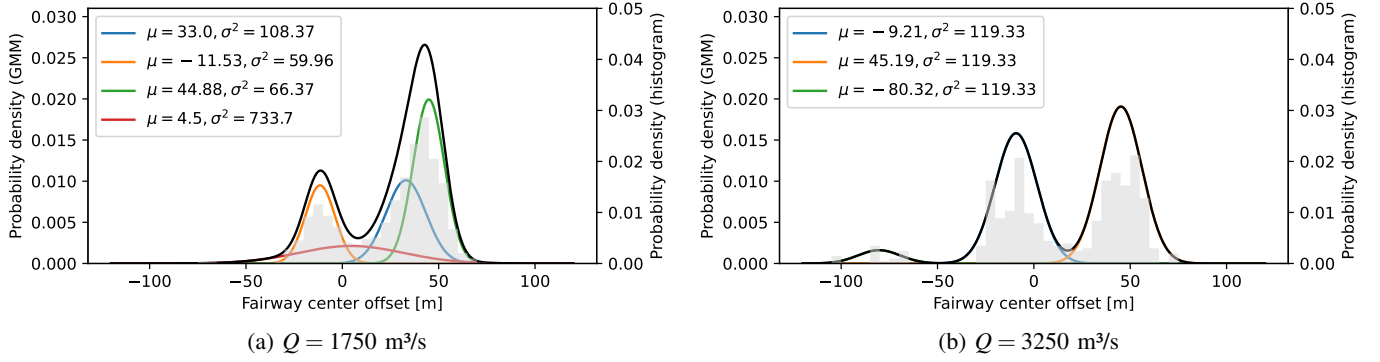


Fig. 1: Example of lateral GMMs: Distribution of fairway center offsets of upstream navigating vessels at Rhine KM 598.6 at different discharges (Q).

probability density curves. By providing these samplings as input, the model is not constrained to a specific cluster. Rather, it incorporates the entire distribution information, ensuring a more comprehensive understanding of vessel behavior.

III. METHODOLOGY

A. Statistical feature generation

The driving behavior of inland vessels navigating free-flowing rivers is significantly affected by the discharge. Discharge Q refers to the volume of water flowing through a cross-section of a waterway per unit of time. It is often given in cubic meters per second. Discharge variation result in modified flow velocities and river depths which affect the vessel's speed and maneuverability [23]. Incorporating discharge information into prediction models is therefore expected to enhance the accuracy of vessel trajectory predictions. Thus, the input data is augmented with feature vectors that characterize the discharge-dependent distribution of the lateral offsets from the fairway center and of the dislocation velocity in terms of waterway kilometer (KM) distance travelled per time interval - both at the vessel's current position as well as in the river section the vessel is heading to. Given the presumed multi-modal nature of both distributions, Gaussian Mixture Models (GMMs) are utilized to approximate them. A GMM can be employed to represent data assumed to originate from a combination of multiple Gaussian distributions. It is a parametric probability density function which returns the probability density by summing up the weighted densities of the Gaussian components. The estimation of the GMM parameters is typically accomplished through the Expectation-Maximization algorithm. To achieve a GMM that best fits the data, two hyperparameters, the number of mixture components and covariance constraints (spherical, diagonal, tied, or full), can be determined using grid search. The model with the lowest Bayesian Information Criterion (BIC) is selected. The BIC is defined as $b \times \log(S) - 2\log(\hat{L})$, where b is the number of parameters of the GMM, S is the dataset size, and \hat{L} is the likelihood of the data under the given model. This evaluation criterion penalizes higher numbers of parameters to avoid overfitting, while maximizing the log-likelihood of

the data. In this study, the maximum number of mixture components used in the grid search is set to 4 which is assumed to be sufficient for the distributions considered. For the representation of the fairway center offset distributions, GMMs are fitted on fairway center offsets of vessel positions from AIS. A separate GMM is generated for each navigation direction, waterway hectometer, and discharge. To avoid data scarcity issues, the KMs of the trajectories are rounded to full hectometers and the measured discharges to the nearest multiple of 250 m³/s. These GMMs are referred to as “lateral GMMs” in the subsequent sections. To model the distributions of KM distances travelled per minute, the waterway is divided into four artificial “lanes”: two inner lanes, which cover the region between fairway center and right and left boundary, respectively, and two outer lanes, encompassing the remaining river area left and right of the fairway. A GMM is fitted on the KM distances per minute of the vessel trajectories from AIS - not only per navigation direction, waterway hectometer, and discharge (rounded), but also per lane. These GMMs are termed “longitudinal GMMs” henceforth. In Fig. 1a and 1b, depicting the probability density curves and components of lateral GMMs for the upstream direction at a selected Rhine KM, the impact of a discharge variation on the fairway center offsets is evident. At a lower discharge, the vessels typically navigate within 50 m to the left of the fairway center (negative offset), while at a higher discharge, they extend their navigation further to the left side. In Fig. 2, the four probability density curves and underlying components of the longitudinal GMMs per lane at a selected Rhine KM and discharge for upstream navigating vessels are shown. From the far left (Fig. 2a) to the far right lane (Fig. 2d), a decrease of the most likely KM distances per minute is visible, and two clearly separable modes are identified by the GMM for the far right lane. Interestingly, shared variances are resulting in the lowest BIC scores for most GMMs depicted in Fig. 1 and Fig. 2. Once the optimal lateral GMMs are obtained for each navigation direction, hectometer, and discharge, bivariate spline interpolation is applied on sampling vectors of the lateral GMMs of each direction and discharge to obtain a function $gmm_{lat}^{dir,q}$, that, for a given navigation direction dir

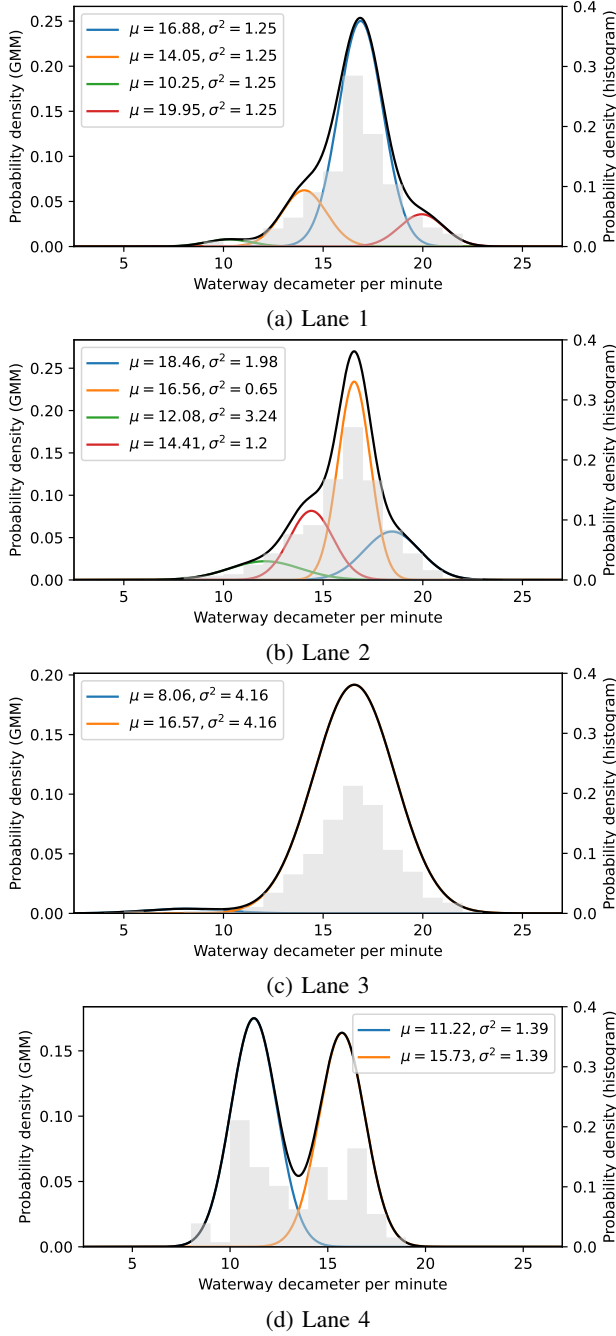


Fig. 2: Examples of longitudinal GMMs: Distributions of KM distance (in decimeters for better visibility) per minute at different locations in relation to the fairway at Rhine KM 608.4 at a discharge of 1000 m³/s (upstream navigation).

and discharge q , returns the probability density value for any requested KM in the covered river section k_{min} to k_{max} and any covered fairway center offset in the range $-m_{max}$ to m_{max} . This interpolation is required as fitting a GMM for every potential KM (given with a precision of centimeters) for the vessel positions from AIS would be impractical due to the resulting data scarcity and computational demands. In Algorithm 1, a lookup dictionary L_{lat} for the considered river section, fairway

Algorithm 1: Generation of lookup dictionary L_{lat} for on-demand generation of vector p_{lat}

Input : Lateral GMMs for directions $\{up, down\}$, discharges $Q = \{q_{min}, q_{min} + 250, \dots, q_{max}\}$, and KMs $K = \{k_{min}, k_{min} + 0.1, \dots, k_{max}\}$; offsets $M = \{-m_{max}, -m_{max} + 1, \dots, m_{max}\}$.

Output: Lookup dictionary L_{lat} containing for each direction dir and $q \in Q$ the function $gmm_{lat}^{dir,q}$.

```

1 Function Interpolate (lateral GMMs,  $Q$ ,  $K$ ,  $M$ ) :
2   Let  $L_{lat}$  be an empty dictionary;
3   foreach  $dir \in \{up, down\}$  do
4     foreach  $q \in Q$  do
5       Let  $A_{lat}[[K]][[M]]$  be an empty 2D array;
6        $j = 0$ ;
7       foreach  $k \in K$  do
8          $\hat{p}_{lat}^{(i)} = \sum_{c=1}^C pdf_c(i - m_{max}; \mu_c, \sigma_c^2)$ ;
          // Sample  $\hat{p}_{lat}$  of length  $|M|$  from
          the lateral GMM for  $q, dir$ , and  $k$ ,
          with  $C$  mixture components
           $\mathcal{N}(\mu_c, \sigma_c^2)$ ,  $i \in \{0, 2m_{max}\}$ .
           $A_{lat}[j] = \hat{p}_{lat}$ ;
           $j = j + 1$ ;
7       end
8        $gmm_{lat}^{dir,q} = \text{BivariateSpline}(K, M, A_{lat})$ ;
          // Interpolate over  $A_{lat}$ , with
          coordinates  $K$  and  $M$ .
9        $L_{lat}[(dir, q)] = gmm_{lat}^{dir,q}$ ;
10    end
11  end
12  return  $L_{lat}$ 

```

center offset range, and discharge range q_{min} to q_{max} is constructed. This dictionary is subsequently employed to obtain on the fly the D -dimensional vector of probability densities, $p_{lat} = (d_{k_t}^1, \dots, d_{k_t}^D)^T$ for the KM k_t of the vessel's position at time step t and D ordered and regularly spaced fairway center offsets. This vector constitutes a sampling vector of the probability density curve of the distribution of fairway center offsets. It is used as one of the additional inputs to the transformer model (Section III-B) to represent the navigation context. The dimension D is a hyperparameter and determined by $2m'r_d + 1$, where $m' \leq m_{max}$ is the maximum absolute fairway center offset chosen and r_d is the selected resolution for the offset. For the longitudinal case, the procedure is identical, except for the additional consideration of the lane, thus resulting in functions designated as $gmm_{lat}^{dir,q,lane}$. The V -dimensional sampling vector $p_{lon} = (v_{k_t}^1, \dots, v_{k_t}^V)^T$ contains sampled densities of the probability density curve representing the distribution of KM distances per minute at the current KM k_t in the lane the vessel is currently located in. Note that the entries both in p_{lat} and p_{lon} are obtained from interpolation and are thus approximate probability density values. Analogously to D , V is a hyperparameter, determined

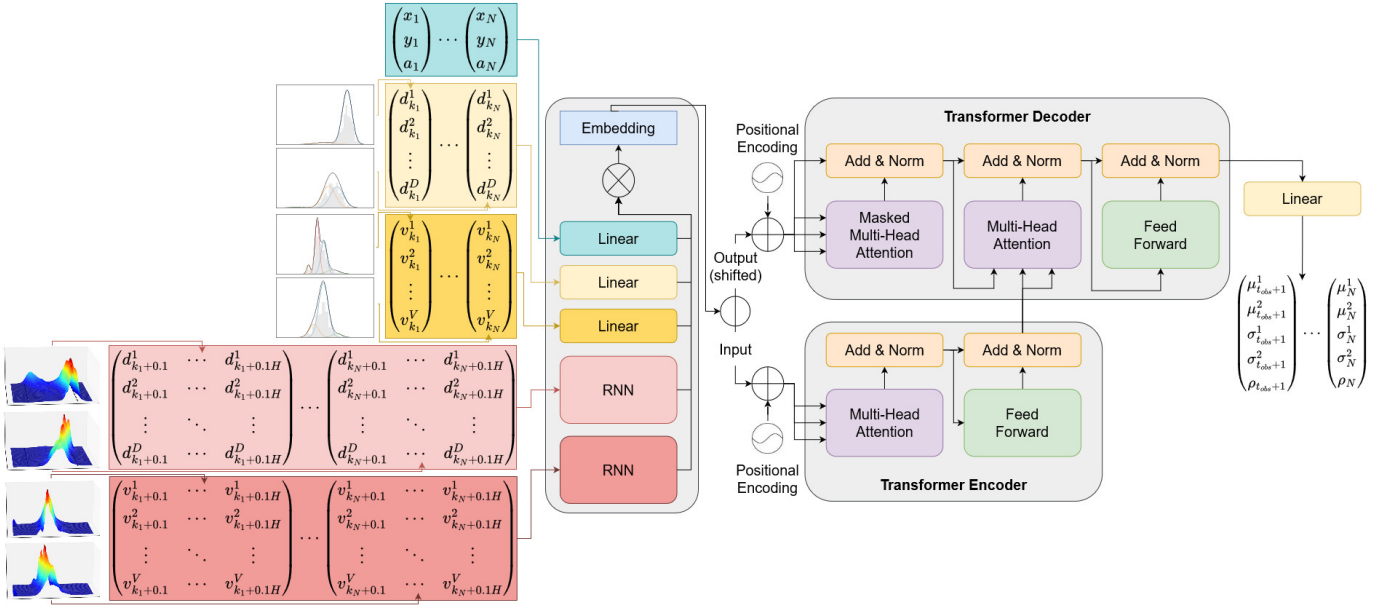


Fig. 3: Architecture of the proposed transformer model for inland vessel trajectory prediction.

by $s_{max}r_v + 1$, where s_{max} is the maximum KM distance per minute that is being considered, and r_v the resolution.

B. Trajectory prediction model

The trajectory prediction task is approached as a time series prediction problem, wherein the vessel's locations and additional contextual data within an observation time interval are used to predict the positions for a subsequent time interval. This is achieved through training a transformer encoder-decoder model to map the input sequence $i_1, \dots, i_{t_{obs}}$ to the output sequence $o_{t_{obs}+1}, \dots, o_N$. Specifically, the input time series element i_t at time step $t \in \{1, \dots, t_{obs}\}$ consists of three vectors and two matrices in this study, which is also visible on the left side of the model architecture diagram in Fig. 3. The first vector is the vessel's (dis)location vector, $(x_t, y_t, a_t)^T$, where x_t , y_t , and a_t are, respectively, the fairway center offset, the KM distance per minute, and the inverse of the fairway center radius, with different signs for right and left curves. The second vector is the p_{lat} vector, which is obtained by looking up $gmm_{lat}^{dir,q}$ in L_{lat} for the given navigation direction dir and discharge q and calling it on the current KM k_t and the D pre-defined fairway center offsets $(-m', -m' + r_d, \dots, m' - r_d, m')$. The third vector p_{lon} is obtained analogously from $gmm_{lon}^{dir,q,lane}(k_t, i)$, $\forall i \in (0, 0 + r_v, \dots, v' - r_v, v')$ (the chosen V KM distances per minute). The two matrices represent the lateral and longitudinal distribution information of the river section, the vessel is heading to, as depicted by the surface plots in Fig. 3. They consist of H stacked vectors obtained similarly to p_{lat} and p_{lon} , respectively, for the next H hectometers of the river section the vessel is heading to, assuming a constant discharge and the continuity of the vessel on the current lane. The vessel (dis)location vector $(x_t, y_t, a_t)^T$, and the vectors p_{lat} and p_{lon} are passed through a linear layer each, as shown in Fig. 3. The matrices representing the

distribution information of the river section ahead are passed each through a recurrent layer (such as an LSTM or GRU). The outputs of the linear and recurrent layers are concatenated (\otimes) to obtain the embedding vectors for each sequence element. The embedded sequence is then split (\oplus) and the observation part for the time steps up to $t_{obs} - 1$ passed to the transformer encoder, while the embedding of the last observation sequence element is passed, together with the remaining time series, i.e. the prediction sequence, to the decoder. Thus, the last observation sequence element is used to initialize the transformer decoder. The decoder input is masked to only provide the known information at each prediction time step as described in [17]. Before being processed by the transformer, positional encodings are added (\oplus) to the sequence elements to provide the model with information of the position of each element in the time series. As $y_t, \forall t \in \{1, \dots, N\}$ is the displacement in KM per minute, (x_t, y_t) do not contain sufficient positional information to eliminate the need for positional encodings here. Similarly to [24], it is assumed, that the lateral and longitudinal trajectory features follow a bivariate normal distribution. Thus, the output at time step $t \in \{t_{obs} + 1, \dots, N\}$, o_t , is a 5D vector containing the means, standard deviations, and correlation of a bivariate Gaussian distribution, $(\mu_t^x, \mu_t^y, \sigma_t^x, \sigma_t^y, \rho_t)$. The transformer model is trained to minimize the negative log-likelihood loss of the true future vessel (dis)location features (x_t, y_t) under the predicted bivariate Gaussian distribution, for all predicted time steps.

IV. DATA

Three sections of the Rhine river with straight, slightly curved, and sharply curved sections are covered in this study (KM 556.5-562, 571-579, and 598.5-608.5). For generating the GMMs, AIS data of the year 2020 are used, and for training the transformer model, AIS data of 2021. Both datasets are

TABLE I: Training duration and model complexity

Model	Time per epoch [s]	Last checkpoint	Trainable parameters
N-CSCT	160	926	745.418
Trans	190	995	80.039
GMM-Trans	320	1000	322.215
GMM-Trans-LSTM	870	977	703.207
GMM-Trans-GRU	790	981	690.407

TABLE II: ADE and FDE after 5 minutes.

Model	ADE	FDE
Const-Acc	41.38±28.75	81.98±61.4
Const-Vel	24.91±16.07	43.48±30.52
N-CSCT [10]	18.67±16.09	30.10±29.65
TP-Baseline	29.25±21.68	43.23±35.58
GMM-Baseline	29.71±22.26	48.30±41.85
Trans	18.03±13.05	29.34±25.53
GMM-Trans	15.14±12.23	23.79±23.73
GMM-Trans-LSTM	14.61±12.10	23.00±23.53
GMM-Trans-GRU	14.54±11.94	22.88±23.35

split into trips. A VTP model for upstream navigating vessels is developed, thus downstream trips are discarded. Trips used for training and testing the model are interpolated to 1 min time intervals using Cubic Hermite spline interpolation and fairway geometries (center curvatures and curve orientations) are added. All trips are labeled with KMs, KM distances per minute, fairway center offsets (negative for positions left of the center), and discharge values. The KMs are determined as in [10]. Discharge values are added from measurements of the gauges covering the region of the vessel positions. The considered river sections do not include transitions between regions covered by different gauges. Stationarities between the gauge and the vessel location are considered neglectable in this study due to the short time horizon and consequently limited KM range. The discharge measurements, available at each 15 min, are provided by the German Federal Waterways and Shipping Agency¹ and the German Federal Institute of Hydrology². The AIS data, trip splitting functionality, fairway geometries, and information of the delimitations of the river sections for each gauge are provided by the BAW. For training the model, the labeled trips are shuffled and split into training, validation, and test dataset in a 80-10-10 ratio, and sequences of N time steps are sampled from each trip without overlapping. The total number of sequences obtained is 264.272. The distribution feature vectors p_{lat} and p_{lon} are dynamically added to the sequences during training/validation/testing by using L_{lat} and L_{lon} . Lookup failures are handled by shifting the requested discharge and/or lane up to a tolerance threshold of ± 500 m³/s and ± 1 lane. Any outliers beyond these thresholds are discarded.

V. RESULTS

The prediction performance of the proposed model is compared to several non DL-based baselines and to the context-

sensitive classification transformer - N-CSCT [10]. Additionally, an ablation study is conducted to assess the advantages of incorporating the proposed features which represent the local and discharge-dependent distributions of both lateral offsets from the fairway center and lane-specific KM distances per minute.

A. Baselines & benchmark

Four baselines are employed, ranging from naive extrapolation methods to more sophisticated approaches incorporating statistical evaluations, including the Gaussian Mixture Models fitted in this study:

- Const-Vel: maintains last observed KM distance per minute and fairway center offset constant during prediction;
- Const-Acc: replaces last observed KM distance per minute in Const-Vel by its change rate;
- TP-Baseline [10]: holds mean deviation of observed KM distances per minute from typical ones and observed average Euclidean distance from the typical route constant during prediction; see [10] for details on the generation of the typical route and typical KM distance per minute;
- GMM-Baseline: holds the distance between the last observed fairway center offset and the closest fairway center offset with the maximum density value in the corresponding p_{lat} and the difference between the last observed KM distance per minute and the closest KM distance per minute with maximum density value in the corresponding p_{lon} constant during prediction.

The N-CSCT [10] is used as a benchmark model. It makes use of statistical pre-evaluation results as well, but instead of appending statistical information to the input feature space, the typical route and typical KM distances per minute, also used by TP-Baseline, are utilized to define the lateral and longitudinal dislocation features of the input and output sequences. Additionally, river curvature and curve orientation are used to initialize the decoder [19]. The N-CSCT is trained, validated, and tested on the same data source as the proposed model to ensure comparability.

In this study, the lateral features (fairway center offset) is not a dislocation feature (different from N-CSCT), while the longitudinal feature (KM distance per minute) is. As the longitudinal dislocation is calculated in the forward direction, the KM of the prediction time step t is obtained from the predicted KM distance per minute of the previous time step $t - 1$. For the first prediction, the KM is consequently already known from the last observed KM distance per minute. To ensure comparability with the baselines and the N-CSCT model, the first predicted KM of these models is replaced by the true one.

B. Model variants of ablation study

Four model variants are examined, each differing in the input features considered and/or the type of recurrent neural network (RNN) utilized (if applicable):

¹<https://www.gdws.wsv.bund.de>

²<https://www.bafg.de>

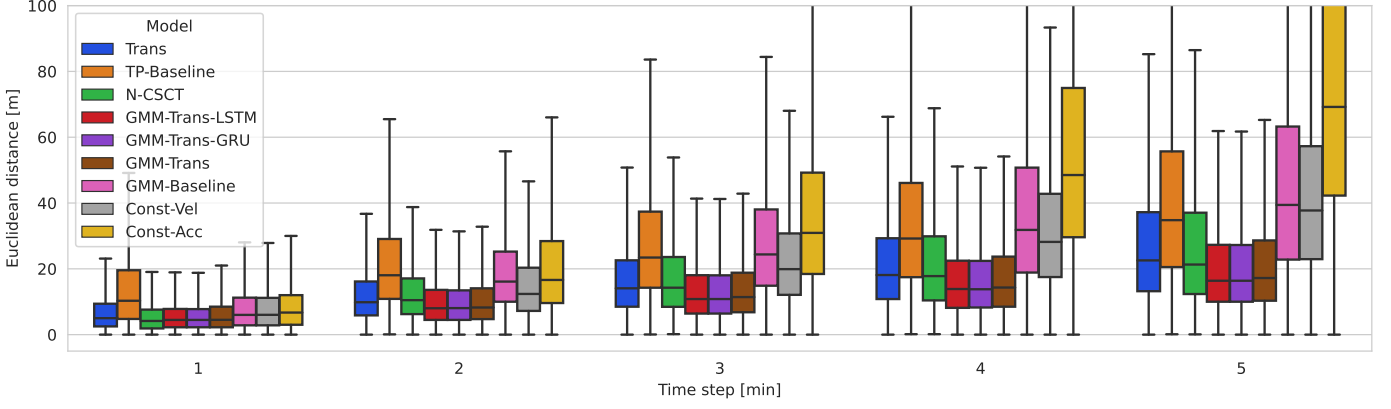


Fig. 4: Prediction error per time step of proposed model variants and baselines.

- Trans: only the vessel (dis)location vectors $(x_t, y_t, a_t)^T$, $t \in \{1, \dots, t_{obs}\}$ are used as input;
- GMM-Trans: p_{lat} and p_{lon} vectors are processed in addition to the vessel (dis)location vectors;
- GMM-Trans-LSTM: considers all proposed input feature vectors and matrices, and the matrices are processed by an LSTM;
- GMM-Trans-GRU: same as GMM-Trans-LSTM but matrix processing is realized with a GRU.

C. Implementation details

All models are trained for up to 1000 epochs with a batch size of 128 and 16 dataloader workers. The training is performed on CPUs of type AMD EPYC 7742 (Rome) with 64 cores and a clock rate of 2.25 GHz. Insight into the training duration and complexity of the proposed model, its variants, and the N-CSCT model are given in Table I. The higher number of trainable parameters in N-CSCT stems from the classification re-framing of the regression task leading to high-dimensional inputs, and a complexer transformer architecture. While the proposed model has one attention head, a single encoder and decoder layer, and a feedforward layer size of 512, the N-CSCT has two attention heads, two encoder and decoder layers, and a feedforward layer size of 1024.

D. Prediction performance

Five time steps (i.e. 5 min) are predicted based on an observation sequence of 6 time steps (the last one used for decoder initialization (see Sec. III-B)). While the N-CSCT processes only 5 time steps as input, they correspond to 6 actual positions since the lateral and longitudinal displacements are given as deltas to the next time step. The prediction error of a sample is measured by the average displacement error (ADE) calculated by taking the average of the Euclidean distances between predicted and actual positions over all predicted time steps, and the final displacement error (FDE), only considering the Euclidean distance between prediction and target at the last prediction time step. Given that the output of the proposed model comprises the distribution parameters of a bivariate

Gaussian distribution rather than a specific position, the final prediction sequence is obtained by sampling 40 pairs of fairway center offsets and KM distances per minute per prediction step from the obtained distribution and taking the mean as the actual prediction. In Table II, the ADE and FDE with standard deviation are reported for all model variants, baselines, and the benchmark model. In Figure 4, the boxplots illustrate the distribution of the FDE across increasing prediction horizons. The best performing model is the GMM-Trans-GRU, suggesting that the integration of all GMM-based features enhances prediction accuracy. While N-CSCT and Trans exhibit similar means, the N-CSCT's standard deviations are notably higher. Surprisingly, the more sophisticated baselines are surpassed by Const-Vel. Furthermore, TP-Baseline outperforms GMM-Baseline, despite relying on a single typical route and a single typical speed per hectometer, overlooking multi-modality and speed variations along the fairway width. Overall, the proposed model and its variants demonstrate superior performance in terms of ADE and FDE for prediction lengths > 1 . For the first prediction time step, N-CSCT performs marginally better, as depicted in Fig. 4.

VI. CONCLUSION

The results of this study show that the integration of GMM-derived feature vectors, which represent local and discharge-dependent distributions of fairway center offsets and waterway kilometer distances per minute, leads to reduced prediction errors in the comparison. The comparative study is characterized by the omission of this distribution information or the exclusive consideration of a single typical route and typical speeds per hectometer, which - as the result demonstrates - oversimplifies the actual complexity of the navigation behavior of inland vessels. Best outcomes are observed when distribution information from both the current vessel location and the upcoming river section are considered. Future works should address several challenges. First, a model architecture and hyperparameter optimization is required. A preliminary hyperparameter test, not detailed in this contribution, showed

that optimization potential exists when increasing the model complexity. Architectural changes, such as consistently using transformers for all sequence encoding tasks or replacing RNNs with CNNs or GNNs to encode the “maps” derived from stacked vectors of lateral and longitudinal distribution information, need to be investigated. Also, the superiority of transformers over RNNs in the given sequence-to-sequence task needs to be demonstrated. Additionally, the scope of discharge data needs to be expanded to include more variations, and variations of the discharge at river confluences need to be taken into account. Furthermore, individual vessel characteristics, such as width, length, and draught, significantly impact the driving behavior and should be considered. As these values are manually added to the AIS data, their inclusion is not trivial, necessitating reliability checks in the data preprocessing stage. Other relevant AIS features that could improve the prediction quality include heading and maneuver indicator. However, these features are often missing due to the absence of required sensors. One major limitation is the unawareness of the traffic situation. Nearby vessels might have a strong influence on the navigation patterns. Thus, the usage of fine-grained spatio-hydrological distribution information in a joint trajectory prediction models such as the one proposed in [20] and [25], which are becoming increasingly relevant especially for the inland domain, is therefore planned. It is expected that the proposed distribution information can help a joint trajectory prediction model to better account for the surrounding vessels’ intents. By addressing these challenges and limitation in future research, the accuracy and robustness of inland vessel trajectory prediction models can be further improved, ultimately enhancing safety and efficiency in navigational decision-making.

ACKNOWLEDGMENT

The authors thank the BAW for the provision of AIS data, river-specific data, and the training infrastructure.

REFERENCES

- [1] Y. Yang, Y. Liu, G. Li, Z. Zhang, and Y. Liu, “Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 183, 2024, 103426.
- [2] H. Li, H. Jiao, and Z. Yang, “Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis,” *Engineering Applications of Artificial Intelligence*, vol. 126, 2023, 107062.
- [3] X. Zhang, X. Fu, Z. Xiao, H. Xu, and Z. Qin, “Vessel Trajectory Prediction in Maritime Transportation: Current Approaches and Beyond,” *IEEE Trans. on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 19 980–19 998, 2022.
- [4] H. Feng, G. Cao, H. Xu, and S. S. Ge, “IS-STGCNN: An Improved Social spatial-temporal graph convolutional neural network for ship trajectory prediction,” *Ocean Engineering*, vol. 266, 2022, 112960.
- [5] J. Sekhon and C. Fleming, “A Spatially and Temporally Attentive Joint Trajectory Prediction Framework for Modeling Vessel Intent,” in *Proc. of the 2nd Conference on Learning for Dynamics and Control*, vol. 120, Online, June 11–12, 2020, pp. 318–327.
- [6] R. W. Liu, M. Liang, J. Nie, Y. Yuan, Z. Xiong, H. Yu, and N. Guizani, “STMGCN: Mobile Edge Computing-Empowered Vessel Trajectory Prediction Using Spatio-Temporal Multigraph Convolutional Network,” *IEEE Trans. on Industrial Informatics*, vol. 18, no. 11, pp. 7977–7987, 2022.

- [7] P. Dijt and P. Mettes, “Trajectory Prediction Network for Future Anticipation of Ships,” in *Proc. of the 2020 Int. Conf. on Multimedia Retrieval*, Dublin, Ireland, June 8–11, 2020, pp. 73–81.
- [8] S. Mehri, A. A. Alesheikh, and A. Basiri, “A Contextual Hybrid Model for Vessel Movement Prediction,” *IEEE Access*, vol. 9, pp. 45 600–45 613, 2021.
- [9] P. Huang, Q. Chen, D. Wang, M. Wang, X. Wu, and X. Huang, “TripleConvTransformer: A deep learning vessel trajectory prediction method fusing discretized meteorological data,” *Frontiers in Environmental Science*, vol. 10, 2022, 1012547.
- [10] K. Donandt, K. Böttger, and D. Söffker, “Improved context-sensitive transformer model for inland vessel trajectory prediction,” in *Proc. of the 26th Int. Conf. on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain, Sep. 24–28, 2023, pp. 5903–5908.
- [11] S. Capobianco, N. Forti, L. M. Millefiori, P. Braca, and P. Willett, “Uncertainty-Aware Recurrent Encoder-Decoder Networks for Vessel Trajectory Prediction,” in *24th IEEE Int. Conf. on Information Fusion (FUSION)*, Sun City, South Africa, Nov. 1–4 2021.
- [12] B. Murray and L. P. Perera, “An AIS-based deep learning framework for regional ship behavior prediction,” *Reliability Engineering & System Safety*, vol. 124, no. 5, p. 107819, 2021.
- [13] D.-D. Nguyen, C. L. Van, and M. I. Ali, “Vessel Trajectory Prediction using Sequence-to-Sequence Models over Spatial Grid,” in *Proc. of the 12th ACM Int. Conf. on Distributed and Event-based Systems*, Hamilton, New Zealand, Jun. 25–29 2018, p. 258–261.
- [14] L. You, S. Xiao, Q. Peng, C. Claramunt, X. Han, Z. Guan, and J. Zhang, “ST-Seq2Seq: A Spatio-Temporal Feature-Optimized Seq2Seq Model for Short-Term Vessel Trajectory Prediction,” *IEEE Access*, vol. 8, pp. 218 565–218 574, 2020.
- [15] N. Forti, L. M. Millefiori, P. Braca, and P. Willett, “Prediction of Vessel Trajectories From AIS Data Via Sequence-To-Sequence Recurrent Neural Networks,” in *Proc. of the 2020 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Online, 2020, pp. 8936–8940.
- [16] R. Jurkus, J. Venskus, and P. Treigys, “Application of coordinate systems for vessel trajectory prediction improvement using a recurrent neural networks,” *Engineering Applications of Artificial Intelligence*, vol. 123, 2023, 106448.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *NIPS’17: Proc. of the 31st Int. Conf. on Neural Information Processing Systems*, Long Beach California, USA, Dec. 4–9, 2017, pp. 6000–6010.
- [18] D. Nguyen and R. Fablet, “A transformer network with sparse augmented data representation and cross entropy loss for ais-based vessel trajectory prediction,” *IEEE Access*, vol. 12, pp. 21 596–21 609, 2024.
- [19] K. Donandt, K. Böttger, and D. Söffker, “Short-term inland vessel trajectory prediction with encoder-decoder models,” in *Proc. of the 25th Int. Conf. on Intelligent Transportation Systems (ITSC)*, Macau, China, Oct. 8–12, 2022, pp. 974–979.
- [20] R. W. Liu, W. Zheng, and M. Liang, “Spatio-temporal multi-graph transformer network for joint prediction of multiple vessel trajectories,” *Engineering Applications of Artificial Intelligence*, vol. 129, 2024, 107625.
- [21] B. Murray and L. P. Perera, “A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data,” *Ocean Engineering*, vol. 209, p. 107478, 2020.
- [22] B. R. Dalsnes, S. Hexeberg, A. L. Flaten, B.-O. H. Eriksen, and E. F. Brekke, “The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-Based Vessel Trajectory Prediction,” in *Proc. of the 21st Int. Conf. on Information Fusion (FUSION)*, Piscataway, NJ, USA, July 10–13, 2018, pp. 580–587.
- [23] M. Heinz, B. Söhngen, S. Doychev, H. Broß, H.-G. Heidenstrecker, A. Hüsiger, C. Meyer-Mölleringhof, and M. Lohbeck, *Driving Dynamics of Inland Vessels - Vessel Behaviour on European Inland Waterways and Waterway Infrastructure with Special Respect to German Waterways*. Karlsruhe, Germany: Bundesanstalt für Wasserbau (BAW), 2016.
- [24] A. Vemula, K. Mueller, and J. Oh, “Social Attention: Modeling Attention in Human Crowds,” in *Proc. of the 2018 IEEE Int. Conference on Robotics and Automation (ICRA)*, Piscataway, NJ, USA, May, 21–26, 2018, pp. 4601–4607.
- [25] S. Wang and Z. He, “A prediction model of vessel trajectory based on generative adversarial network,” *Journal of Navigation*, vol. 74, no. 5, pp. 1161–1171, 2021.