

A Deep Learning Approach for Location Independent Throughput Prediction

Josef Schmid, Mathias Schneider and Alfred Höß
Ostbayerische Technische Hochschule
Amberg-Weiden
92224 Amberg, Germany
Email: {j.schmid, mat.schneider, a.hoess}@oth-aw.de

Björn Schuller
ZD.B Chair of Embedded Intelligence
for Health Care and Wellbeing,
University of Augsburg, 86159 Augsburg, Germany
Email: schuller@ieee.org

Abstract—Mobile communication has become a part of everyday life and is considered to support reliability and safety in traffic use cases such as conditionally automated driving. Nevertheless, prediction of Quality of Service parameters, particularly throughput, is still a challenging task while on the move. Whereas most approaches in this research field rely on historical data measurements, mapped to the corresponding coordinates in the area of interest, this paper proposes a throughput prediction method that focuses on a location independent approach. In order to compensate the missing positioning information, mainly used for spatial clustering, our model uses low-level mobile network parameters, improved by additional feature engineering to retrieve abstracted location information, e.g., surrounding building size and street type. Thus, the major advantage of our method is the applicability to new regions without the prerequisite of conducting an extensive measurement campaign in advance. Therefore, we embed analysis results for underlying temporal relations in the design of different deep neuronal network types. Finally, model performances are evaluated and compared to traditional models, such as the support vector or random forest regression, which were harnessed in previous investigations.

Index Terms—throughput prediction, deep learning, cellular network

I. INTRODUCTION

Mobile networks are evolving with a high pace in the recent years and every released cellular network generation provides a higher amount of bandwidth as well as lower latency. This technological change further enables novel services and finally the needs and expectations of the end-user. For example, on the one hand transport industry enhances their vehicles by implementing safety features incorporating highly accurate data from a centralized server. Whereas on the other hand, passengers demand to use the network capability for their private needs at the same time, i.e. media streaming. Since overall bandwidth is a limited resource depending on the cellular network coverage at the location, exceeding this maximal capacity is a probable consequence. This effect especially occurs at transportation hubs where the traffic is very dense. This was shown in our previous analysis on a motorway around Frankfurt am Main in Germany [1]. One possible improvement for these situations is the prediction of the available throughput during driving, utilized either to adjust the bandwidth allocation based on messages' priorities or to preload data when the capacity is available.

In contrast to previous investigations, which contemplated location reliant methods for prediction, this paper investigates the performance of location independent time series based deep learning models. This approach is chosen to overcome preliminary measure campaigns at unmapped areas. Thereby, we examine different Deep Neural Networks (DNNs), including Feedforward Neural Networks (FFNs), with a special focus on Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) models. The latter are according to recent research even more suitable for time-related regression problems. These deep learning approaches are compared to traditional machine learning methods such as Support Vector Regression (SVR) or Random Forest Regressor (RFR) models which proved to achieve good results in our recent work [2].

The rest of this paper is structured as follows. Section II presents related work regarding throughput prediction in vehicular context. Subsequently, Section III introduces two geographically distinct datasets utilized in this work and explains the separation for training and valuation in detail. Since dealing with time series data, Section IV provides an analysis of the temporal characteristics of the selected features to improve the design of the deep learning models. Section V introduces the actual models and their topology. Thereafter, error metrics are described and final results are gathered in Section VI comprising an evaluation with regards to the two considered test tracks. Finally, a conclusion of this work is given in Section VII.

II. RELATED WORK

Recent algorithms in the research field of throughput prediction in cellular networks are mainly categorized into geographical and time series based approaches. Especially for the latter, the usage of Machine Learning (ML) became the focus of attention in the recent years. Nevertheless, due to the wide possibilities how to apply these models, the strategies between the authors vary, and dominant ones will be introduced in the next paragraphs.

As a first reference, Yue et al. [3] leveraged a Random Forest (RF) based approach to predict the throughput of Long-Term Evolution (LTE) connections. Thereby, their prediction model utilized different low level LTE parameters, i.e. Reference Signal Receiving Power (RSRP) and Reference Signal Received

Dataset	Date	Datapoints	Rounds	Distance [km]	$\sigma_{TP_{DL}}$ [kB/s]	$\mu_{TP_{DL}}$ [kB/s]
<hr/>						
Amberg (training)		329 518	60	1626.68	1281.75	996.67
	03.04.2018 (Tue.)	52 621	9	243.72	1412.28	1032.96
	04.04.2018 (Wed.)	21 716	4	108.99	1505.17	927.84
	05.04.2018 (Thu.)	122 710	23	626.47	1153.49	951.57
	22.06.2018 (Fri.)	132 471	24	647.50	1296.98	1042.28
<hr/>						
Amberg (evaluation)		21 565	4	110.16	946.64	948.88
	03.04.2018 (Tue.)	4891	1	27.66	715.81	848.10
	04.04.2018 (Wed.)	5678	1	27.57	786.84	852.23
	05.04.2018 (Thu.)	4883	1	27.68	1357.16	1104.78
	22.06.2018 (Fri.)	6113	1	27.25	799.88	1000.12
<hr/>						
Aschaffenburg (evaluation)		49 804	9	165.32	901.61	865.34
	04.06.2018 (Mon.)	22 865	3	58.38	610.95	674.17
	05.06.2018 (Tue.)	12 464	3	56.73	969.94	848.00
	06.06.2018 (Wed.)	14 475	3	50.21	1127.15	1207.34
<hr/>						

Table I: Properties of a measurement campaign in Amberg and Aschaffenburg, separated in training and evaluation datasets.

Quality (RSRQ). Other investigations conducted by Ghasemi kept the idea of a RF based model, but used the parameters collected with the application OpenSignal¹ to predict LTE down- and uplinks [4]. The RF algorithm also attained decent results in our previews work [2]. Thereby, it was compared to the performance of other traditional ML methods such as Linear Regression (LR) and achieved the same order of accuracy as SVR models.

On the other side, there are models incorporating the geographical coordinate of the measurement to anticipate the throughput. One approach, presented by Tanni et al. in [5], collected the data of public transportation bus drives and applied a geostatistical technique called Kriging to interpolate the throughput measurements. But also in the area of geo-based models machine learning is gaining popularity, resulting in hybrid approaches. In [6], Sliwa et al. built a multilayer map consisting of RSRP, RSRQ, Signal-to-Noise-Ratio (SNR) and Channel Quality Indicator (CQI). With this information, they use an M5 Decision Tree model to optimize the overall system performance and provide interference-free coexistence with human data traffic that is using the same public cellular network. Another technique was introduced by Wei et al. [7]. They inserted the geological coordinate directly as input for the LSTM RNN models. Accordingly, this procedure constitutes a geo-based model without explicitly modeling a map, but implicitly utilizes position information in the DNN model. The introduced geo-based approaches result in two disadvantages during model inference. Firstly, they require an accurate estimate of the forthcoming position, which itself refers to the problem of trajectory prediction mentioned in [8]. Secondly, this method still relies on the existence of historical data to build its model, which is one major preliminary of most geo-based models.

Hence, this paper focuses to design a location independent time series based model to compensate these disadvantages.

Therefore, a modern Neural Networks (NN) model, the LSTM RNN model is taken into consideration, since it is leveraged in related throughput prediction problems [9], [10]. This paper compares its performance and ability to reflect underlying temporal correlations during driving and evaluates them with respect to our former traditional ML models.

III. DATASETS

For the evaluation of the deep learning models' location, independent datasets at two locations are leveraged. The first dataset comprises a new measurement campaign at a test area used in [2], [11]. It is located in the vicinity of the city of Amberg, Bavaria, in Germany. As described in the previous work, this test round is selected, because of its variety of traffic scenarios, such as urban, interurban, and motorway roads. In comparison with the former Amberg dataset [2], [11], the measurement setup is changed to enable the Multiple-Input and Multiple-Output (MIMO) capability of the used Qualcomm modems. Besides this modification, the setup remained the same, including the TCPAnalyzer tool chain for conducting the measurements with a sampling frequency of 3.33 Hz. Nevertheless, leveraging MIMO results in higher throughput rates compared to the former dataset that consequently increases the complexity for prediction models.

In addition, a track with the same road types is chosen for the evaluation and includes a route between Dudenhofen and Aschaffenburg, Hessen, in Germany as depicted in Figure 1.

Table I provides a detailed overview of the datasets and the partitioning. Thereby, the overall throughput metrics for each set, comprising the arithmetic mean and standard deviation, are within the same order of magnitude for these test sets, although they differentiate in the number of samples.

The reason for defining and acquiring two distinct datasets is to apply the following gradual strategy. Whereas the Amberg measurements are used for the training and a first location dependent evaluation of the models, the data points acquired in Aschaffenburg are exclusively harnessed to determine the

¹<http://www.opensignal.com>



Figure 1: Test area between Aschaffenburg and Dudenhofen in Hessen, Germany, including road type annotation.

model inference performance for unseen inputs at an unknown place. This partitioning will serve as a basis to evaluate the location independence in Section VI.

IV. TIME CHARACTERISTICS OF THE USED DATA FEATURES

In our previous work [2], we intensively examined the correlation between throughput and other network parameters as part the feature selection process prior to the model training. Since LSTM RNN models further involve temporal relations, which were not explicitly explored, we extend this analysis. The objective for this detailed investigation is to determine the optimal memory length of the LSTM RNN, also referred to as lag. Setting the input memory length has crucial effects on the training and model inference. Expanding the input time series for prediction causes that less data can be used for training, and more time elapses before the first prediction can be computed.

Thereby, the widely used autocorrelation function is utilized to elaborate the correlations in sequential data. Based on the downsampled time series, in which each data point aggregates the measurements over a sliding window of 15 s, these autocorrelation values are calculated. Afterwards, the number of relevant consecutive lags is determined by solely considering the number of values within a confidence interval of 95 %. This procedure is applied for every selected feature in all training set rounds. All lag results are summarized in the box plot depicted in Figure 2. The box plot shows that the median of autocorrelation values for all features is covered in a lag of four, corresponding to an overall measurement interval of one minute. Especially for important features such as RSRP, RSRQ, Reference Signal Strength Indicator (RSSI), Round-Trip Time (RTT), and the throughput, this determined lag, indicated by the dashed line, is not exceeded by any value. A detailed description of all features can be found in [2]. Accordingly, this memory length is used to prepare the training dataset for the LSTM RNN models.

V. DEEP LEARNING PREDICTION METHODS

In the following section, models used for training location independent prediction are introduced. Thereby, our investigations are focused on two types of DNN. On the one hand, DNNs with multiple fully connected hidden forward layers, also referred to as FFNs, are considered. On the other hand, LSTM RNNs, which are a more resilient sub-type of RNNs

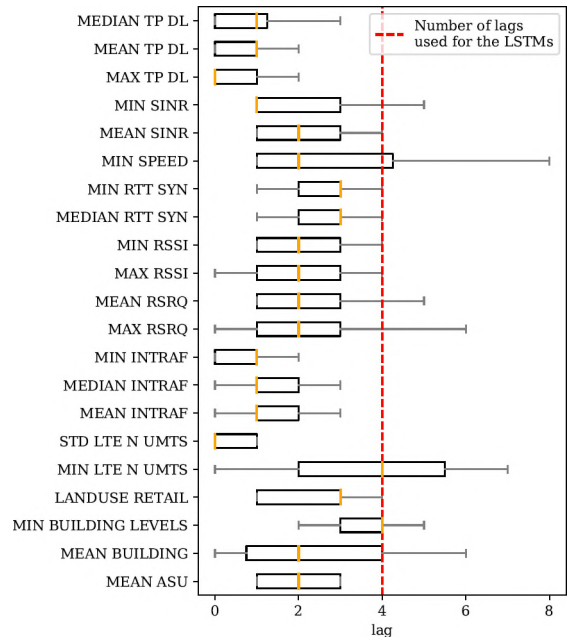


Figure 2: Autocorrelation function values of the training dataset for each selected feature, whereby a dashed line indicates the lags used to prepare the data for RNNs.

against the vanishing gradient problem, are harnessed. They differ from usual DNNs by consisting of neurons with internal states that are capable of keeping track of relations in the input sequence by using an additional feedback connection. Thus, they are convenient for optimizing time series regression problems.

Similar to previous approaches, all models are trained to perform a next step prediction. Because the time series are preprocessed to aggregate measurements in slices, the anticipated result is 15 s in the future. Furthermore, the models are trained to predict the difference from the current value, instead of forecasting the upcoming absolute throughput. This operation must be reversed to obtain the absolute values, which is necessary for the interpretation in the evaluation.

For the FFNs, a topology optimization is carried out. This comprises to examine the effects on the performance of different amounts of hidden layers in combination with a varying number of neurons per layer. As a result, a model comprising six wide hidden layers outperforms all other variation and is used for the detailed evaluation. A schema of the model is shown in Figure 3a.

As an additional preparation for the LSTM RNN models, batches are built from the training datasets, each consisting of ten samples. Similar to the FFN training, various widths and depths of the network structure are tested, starting from a plain Vanilla RNN. In addition, the usage of bidirectional layers is evaluated. The performance result indicates that a unidirectional LSTM RNN model with two stacked layers, depicted in Figure 3b, provides the best outcome.

Regardless of the DNN type, model training is implemented

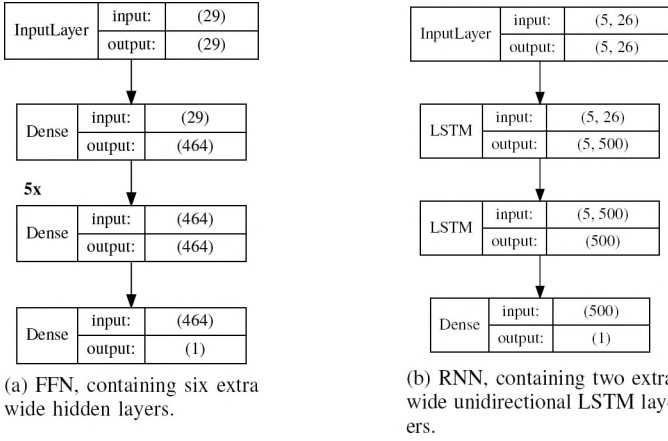


Figure 3: Selected DNN models used in the evaluation.

to utilize 90% of the input data for training and 10% for validation, whereby the latter is used to monitor the training progress of single epochs. As an optimizer, Adam is set with its default Keras parameters and Mean Squared Error (MSE) is used as a loss function. With this setup, a training of 250 epochs is performed. Consequently, all DNNs are trained with regard to this error function and thus it is our main error metric for the location independent evaluation.

VI. COMPARISON OF AI PREDICTION METHODS

As mentioned in Section III, the evaluation of the different prediction models is conducted on unseen data of the Amberg track as well as with the whole data of the Aschaffenburg measurements. Although the models are optimized with regards to the MSE loss function, for the purpose of error calculation, several following functions are leveraged. In these formulas, R_i is defined as the i -th measured value, whereas \hat{R}_i is the i -th forecast:

- **Mean Absolute Error (MAE)** calculates the mean of the absolute difference between real and predicted values.

$$MAE = \frac{1}{n} \sum_{i=0}^n |R_i - \hat{R}_i| \quad (1)$$

- **Mean Relative Error (MRE)** computes the relative error with regards to the minimum of the real and predicted values. Thus, a larger absolute error for higher values is less impactful than the same absolute difference for overall smaller values.

$$MRE = \frac{1}{n} \sum_{i=0}^n \left(\frac{R_i - \hat{R}_i}{\min(R_i, \hat{R}_i)} \right) \text{ s. t. } R_i, \hat{R}_i > 0 \quad (2)$$

- **Mean Squared Error (MSE)** applies the square of the absolute error and accordingly punishes outliers harsher than MAE.

$$MSE = \frac{1}{n} \sum_{i=0}^n \left(\hat{R}_i - R_i \right)^2 \quad (3)$$

- **Mean Squared Relative Error (MSRE)** is a common error, which is highly sensitive even to single outliers. Since it is widespread in the area of throughput prediction, we complement our evaluation to enable comparisons with future model approaches.

$$MSRE = \frac{1}{n} \sum_{i=0}^n \left(\frac{R_i - \hat{R}_i}{R_i} \right)^2 \text{ s. t. } R_i > 0 \quad (4)$$

As denoted in Equation 2 and 4, some error functions do not take all values into account to avoid a possible zero division.

The results of the evaluation for Amberg using all error functions are summarized in Table II. In addition to the deep learning and traditional ML methods, also the metrics for the Persistence Algorithm (PA) are shown. This model is often employed as a baseline for time series regression and uses the current value to determine the prediction for the next step, thus expressed by $\hat{R}_{i+1} = R_i$.

Table II: Error results for the valuation rounds of the Amberg dataset.

Model	MAE [kB/s]	MRE	MSE	MSRE
PA	352.10	0.68	385269.67	3.42
SVR	323.34	0.72	319091.62	7.62
RFR	323.53	0.70	263377.85	6.60
FFN	320.51	0.87	262531.51	21.53
LSTM RNN	305.98	0.76	250693.95	13.76

The evaluation of the data in Amberg shows that all deep learning models are feasible for a prediction and provide a better outcome compared to the PA in terms of MSE. The FFN performs better than the SVR and RFR as well and is solely outperformed by the LSTM RNN.

Furthermore, the DNN methods, trained on the Amberg training dataset, are tested against measurements from the different track in Aschaffenburg to prove that these methods work location independently, without the requirement to feed additional data into the model. Despite these errors metrics, depicted in Table III, are slightly higher for this distinct course, the order of magnitude for error remains the same. Only the difference in the MSRE indicates that these results might contain more and/or severe outlier. The evaluation reveals that with regards to MSE, the LSTM RNN models performs better than the FFN does.

Table III: The Error results for the valuation round of the Aschaffenburg dataset.

Model	MAE [kB/s]	MRE	MSE	MSRE
FFN	368.72	0.86	278730.90	5.66
LSTM RNN	386.19	1.36	269187.09	25.91

Concluding the evaluation, Figure 4 presents a measured throughput plot in Amberg. Moreover, the predicted sequence of the LSTM RNN is additionally plotted into the figure for

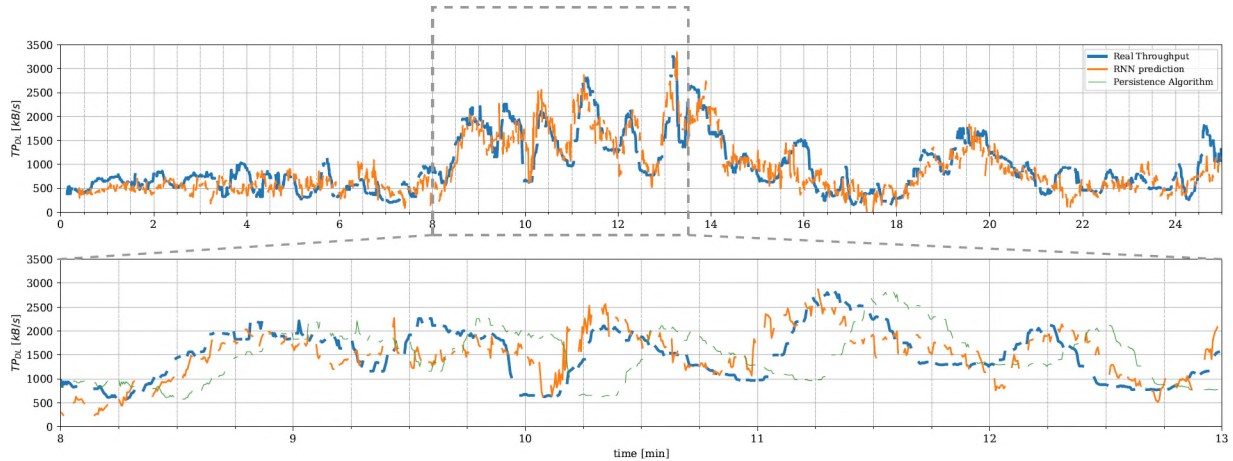


Figure 4: Evaluation of downlink throughput for a round in Amberg including an enlarged segment. Gaps in the sequence are due to filtering mechanisms, i. e., TCP slow-start phase removal. The enlarged plot demonstrates that the LSTM RNN prediction outperforms the Persistence Algorithm, since it reacts more accurate to abrupt changes.

comparison. The enlarged segment, recorded on the motorway, emphasizes the major advantage of DNN over the modest PA models, which is the fast adaption to the dynamic environment. As far as our measurements show, this characteristic is mandatory for an accurate throughput prediction in cellular networks during driving.

VII. CONCLUSION

Mobile network throughput prediction in moving vehicles is an important aspect, when it comes down to optimize the communication in vehicle to network use cases. In this paper, we presented different DNN approaches which allow to apply a decent prediction of the throughput as well as to apply it in an area where no mobile network datasets are previously collected. Our analysis of the time dependency of the feature shows that previous values effect the prediction, which supports the recommendation of various authors to consider LSTM RNN models. Nevertheless, it turns out that the performance still depends on the underlying error and loss functions that are leveraged to optimize the DNNs. In the given scenario, which uses the MSE as main criterion, the LSTM RNN outperforms all other used algorithms. Consequently, this method is also worth considering, if another error or similar scenarios such as prediction of latency. As this parameter is, depending on the application, at least as important for the quality of the communication, we aim to investigate it in our further work. An investigation of the independence of the providers could also be a further starting point to improve the results shown here.

VIII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the following H2020 - ECSEL Joint Undertaking projects and agencies for financial support including funding by the German Federal Ministry for Education and Research (BMBF): PRYSTINE (Grant agreement No. 783190, funding code 16ESE0330).

REFERENCES

- [1] F. Jomrich, J. Schmid, S. Knapp, A. Höß, R. Steinmetz and B. Schuller, "Analysing communication requirements for crowd sourced backend generation of HD Maps used in automated driving," in *2018 IEEE Vehicular Networking Conference (VNC)*, Taipei, Taiwan, 2018, pp. 1-8.
- [2] J. Schmid, M. Schneider, A. Höß and B. Schuller, "A comparison of AI based throughput prediction for moving mobile networks," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 2019, pp. 471-476.
- [3] C. Yue, R. Jin, K. Suh, Y. Qin, B. Wang and W. Wei, "LinkForecast: Cellular Link Bandwidth Prediction in LTE Networks," in *IEEE Transactions on Mobile Computing*, vol. 17, no. 7, pp. 1582-1594, 2018.
- [4] A. Ghasemi, "Predictive Modeling of LTE User Throughput Via Crowd-Sourced Mobile Spectrum Data," in *2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Seoul, South Korea, 2018, pp. 1-5.
- [5] B. Taani and R. Zimmermann, "Spatio-temporal analysis of bandwidth maps for geo-predictive video streaming in mobile environments," in *Proceedings of the 2016 ACM on Multimedia Conference*, Amsterdam, Netherlands, 2016, pp. 888-897.
- [6] B. Sliwa, R. Falkenberg, T. Liebig, J. Pillmann and C. Wietfeld, "Machine learning based context-predictive car-to-cloud communication using multi-layer connectivity maps for upcoming 5G networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, IL, USA, 2018, pp. 1-7.
- [7] B. Wei, W. Kawakami, K. Kanai, J. Katto and S. Wang, "TRUST: A TCP Throughput Prediction Method in Mobile Networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, 2018, pp. 1-6.
- [8] W. Zhang, Y. Liu, T. Liu and C. Yang, "Trajectory prediction with recurrent neural networks for predictive resource allocation," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, Beijing, China, 2018, pp. 634-639.
- [9] S. Xiao and C. Wen, "Dynamic Allocation of 5G Transport Network Slice Bandwidth Based on LSTM Traffic Prediction," in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2018, pp. 735-739.
- [10] R. Madan and P. Mangipudi, "Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN," in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, Noida, India, 2018, pp. 1-5.
- [11] J. Schmid, P. Heß, A. Höß and B. W. Schuller, "Passive monitoring and geo-based prediction of mobile network vehicle-to-server communication," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Limassol, Cyprus, 2018, pp. 1483-1488.