# Application of Ensemble Regression Methods in Elastic Optical Network Optimization

**Aleksandra Knapińska**[1,*][0000−0003−2654−4893],
**Robert Kanimba**[1], **Yusuf Yeşilyurt**[1],
**Piotr Lechowicz**[1,2][0000−0003−2555−5187],
**Krzysztof Walkowiak**[1][0000−0003−1686−3110]

[1] *Wrocław University of Science and Technology, Wrocław, Poland*
[2] *Chalmers University of Technology, Gothenburg, Sweden*
[*] *aleksandra.knapinska@pwr.edu.pl*

**Abstract.** *In this paper, we tackle the problem of estimating the operation of resource allocation algorithms in multilayer optical networks. We show that it is possible to create a regression model simulating a routing and spectrum allocation algorithm to predict four different metrics (i.e., highest occupied slot, average highest occupied slot, sum of occupied slots, and number of active transceivers), only having the input set of connection requests and no information about the underlying topology. We analyze the performance of various ensemble methods, including XGBoost, Random Forest, and stacking models on two large topologies, and demonstrate their good prediction capabilities.*

**Keywords:** *multilayer network, machine learning, metric estimation*

## 1. Introduction

Backbone optical networks are the basis of today's internet communication. Enormous amounts of data are transmitted every second to provide connectivity between even the most distant locations. As we approach capacity limits with the ever-increasing traffic volumes, new traffic-driven allocation algorithms are getting attention in the research community [1, 2] to provide reliable systems, maximizing the amount of provisioned traffic and network stability.

Broad testing is always required to ensure the applicability of the proposed methods to various traffic conditions. Multiple time-consuming simulations are needed to cover multiple possible traffic scenarios. Unfortunately, the scale of the resource allocation problem in backbone networks is immense. In this regard, estimation methods based on machine learning (ML) come in handy to predict the likely outcome without actually solving it for particular instances, which saves excessive time and resources, aligning with green networking paradigms. In particular, methods for estimating various metrics such as quality of transmission [3], latency [4], or bandwidth blocking [5] are continuously developed. However, additional detailed information, such as the number and position of active lightpaths, is usually required for the successful deployment of such models.

In this paper, we tackle the problem of estimating the operation of resource allocation algorithms in multilayer optical networks. We show, for the first time, that it is possible to create a regression model simulating a static routing and spectrum allocation (RSA) algorithm to predict four different metrics (i.e., highest occupied slot, average highest occupied slot, sum of occupied slots, and number of active transceivers), only having the input set of connection requests and no information about the underlying topology. To this end, we analyze the performance of various ensemble methods, including XGBoost, Random Forest, and stacking models on two large topologies and demonstrate their good prediction capabilities.

## 2. Allocation algorithm and network metrics

In this Section, we describe our routing and spectrum allocation algorithm for multilayer backbone optical networks. Given a set of connection requests, each characterized by its source node, destination node, and bitrate, the goal is to allocate all of them in the network to minimize the resources used. We consider a multilayer network model (see Fig. 1) that consists of a physical network topology (optical (EON) layer) and a virtual topology of lightpaths (packet (IP) layer).

The algorithm operates according to the procedure provided in Alg. 1. First, the connection requests are sorted by bitrate (line 1) and then processed one by one (line 2). For each request, the algorithm first checks if there is an existing lightpath between its source and destination with enough spare bandwidth (line 3). If so, it performs *traffic grooming*, i.e., provisions an additional request in an existing lightpath using its free space (line 4). Otherwise, the algorithm sends the lightpath creation request to the optical layer (line 6), providing information about the source and destination node along with the requested bitrate. After the new lightpath is set up, the connection request is allocated in it (line 7).

RSA algorithms can be assessed using various metrics, describing different as-

Algorithm 1: RSA in a multilayer IP-over-EON network.

Sort requests by bitrate
**for** each request **do**
    **if** a direct lightpath from its source to its destination exists and has enough free space **then**
        groom the request into this lightpath
    **else**
        set up a new lightpath in EON layer
        allocate the request into the newly-created lightpath
    **end if**
**end for**

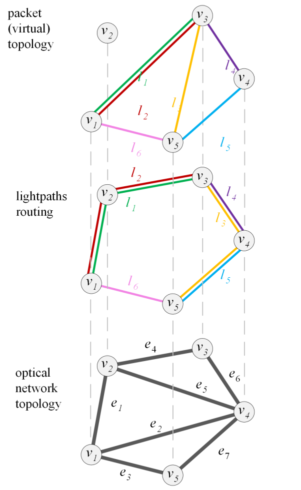


Figure 1: Multilayer IP-over-EON network model.

pects of their operation. In particular, the optical fibers are divided into frequency slots, and the lightpaths are usually provisioned using the distance-adaptive transmission rule [6] – using the most spectrally efficient modulation format for the requested bandwidth considering the path length. Thus, the *highest occupied slot* is a general metric to be minimized, assessing the overall spectrum occupancy. The *average highest occupied slot* describes the spectrum occupancy on network links and gives a broader idea of the network saturation. Additionally, the *sum of occupied slots* gives an idea of the actual utilization of network links considering spectrum fragmentation. Finally, the *number of active transceivers* (devices transmitting and receiving optical signals) can be translated into the network's energy efficiency. For more information about the traffic model and resource allocation with traffic grooming, we refer to [7].

## 3. ML methods for network metric prediction

In this Section, we propose a regression model for estimating the operation of our RSA algorithm and predicting the values of metrics without actually solving the problem. In more detail, we wish to create a *black-box* representation of the developed RSA algorithm that, for a given set of connection requests, would instantly provide the prediction of four metrics (*highest occupied slot*, *average highest occupied slot*, *sum of occupied slots*, and *number of active transceivers*). That way, an operator can expect how a network operated using this algorithm would act in new traffic conditions without conducting time-consuming simulations.

In this work, we explore the applicability of ensemble methods to create the black-box regression model. Specifically, we first evaluate the performance of Random Forest and XGBoost. Then, we combine their strengths through *stacking* [8] and assess the benefits of this hybrid approach. Stacking is an ensemble learning technique that leverages multiple base models and a meta-learner to make predictions. To this end, we use Ridge Regression as a lightweight but capable predictor.

# 4. Numerical experiments

We evaluate the performance of our regression models on two large network topologies (see Fig. 2). For each topology, we generate 100 sets of requests uniformly distributed between node pairs and allocate them in the network using the algorithm described in Sect. 2 to obtain the true metric values. Next, we evaluate all regression models using 5x3 cross-validation. We repeat the experiments for all metrics mentioned in Sect. 2. In other words, we train each model four times so that it learns to predict different metrics from the same inputs.

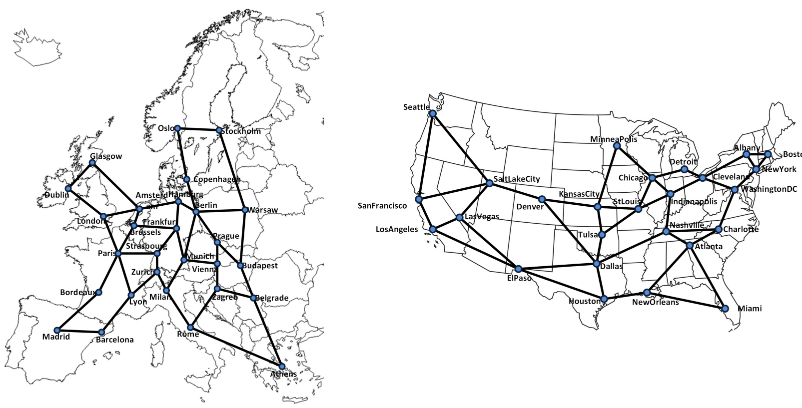


Figure 2: Considered network topologies: Euro28 with 28 nodes and 82 links (left) and US26 with 26 nodes and 84 links (right).

In Tab. 1, we present the results obtained in our experiments. To assess the prediction quality, we use the mean absolute percentage error (MAPE), which allows for a direct comparison of values in different ranges. The first observation is that the prediction quality is satisfactory overall but differs significantly between the network metrics. The highest errors were noted for the *highest occupied slot* and *average highest occupied slot*, while the lowest – for the *number of active* *transceivers* and *sum of occupied slots*. The possible reason might lie within the distribution of these metrics – for the same traffic load, the values of the latter metrics are pretty stable, while the former is more nuanced. In more detail, the number of lightpaths needed to provision specific amounts of traffic is relatively stable and corresponds directly to the number of active transceivers and the sum of occupied slots. The remaining metrics, on the other hand, are more dependent on network fragmentation and, thus, differ more significantly between individual cases of traffic conditions.

Table 1: MAPE for the considered ML models and network metrics.

| | **highest occupied slot** | **avg. highest occupied slot** | **sum of occupied slots** | **n. active transceivers** |
|---|---|---|---|---|
| | | **Euro28** | | |
| XGBoost | 0.1010 | 0.0798 | 0.0491 | 0.0209 |
| Random Forest | **0.0871** | 0.0699 | **0.0448** | **0.0195** |
| Stacked | 0.0881 | **0.0690** | 0.0456 | 0.0199 |
| | | **US26** | | |
| XGBoost | 0.1433 | 0.0986 | 0.0590 | 0.0215 |
| Random Forest | 0.1368 | **0.0856** | **0.0532** | 0.0196 |
| Stacked | **0.1301** | 0.0873 | 0.0554 | **0.0194** |

Between the ML models, the results obtained by Random Forest and the Stacking approach are very comparable, while the XGBoost model noted slightly higher prediction errors. Considering the much higher complexity of the Stacking model, we then recommend employing the Random Forest model as a good balance between prediction quality and computational complexity. Comparing the results between the tested topologies, we do not notice significant differences in the identified trends, both regarding the performance differences between network metrics and between ML models, ensuring the versatility of the obtained results.

## 5. Conclusions

In this paper, we proposed a regression model that estimates the operation of a routing and spectrum allocation algorithm to predict a metric value without actually solving the problem. We showed, how only having the input set of connection requests, it is possible to accurately predict the values of four different metrics (*highest occupied slot*, *average highest occupied slot*, *sum of occupied slots*, and *number of active transceivers*) without the knowledge of the underlying network topology. Through experiments, we demonstrated good predictive capabilities of ensemble methods: XGBoost and Random Forest, and a stacking model combining their strengths.

In the future, we plan to build a multioutput regression model for simultaneous prediction of multiple metrics and employ explainable artificial intelligence techniques to gain insights into the operation of the proposed black-box models.

## Acknowledgment

This work was supported by the National Science Center, Poland under Grant 2019/35/B/ST7/04272.

## References

[1] Panayiotou, T. et al. Survey on machine learning for traffic-driven service provisioning in optical networks. *IEEE Communications Surveys & Tutorials*, 25(2):1412–1443, 2023.

[2] Musumeci, F. et al. An overview on application of machine learning techniques in optical networks. *IEEE Communications Surveys & Tutorials*, 21(2):1383–1408, 2019.

[3] Ibrahimi, M. et al. Machine learning methods for quality-of-transmission estimation. In *Machine Learning for Future Fiber-Optic Communication Systems*, pages 189–224. Elsevier, 2022.

[4] Klinkowski, M. et al. Application of linear regression in latency estimation in packet-switched 5G xHaul networks. In *23rd International Conference on Transparent Optical Networks (ICTON)*. 2023.

[5] Ksieniewicz, P. et al. Pattern recognition model to aid the optimization of dynamic spectrally-spatially flexible optical networks. In *20th International Conference on Computational Science (ICCS)*. 2020.

[6] Jinno, M. et al. Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network. *IEEE Communications Magazine*, 48(8):138–145, 2010.

[7] Knapińska, A. et al. On advantages of traffic prediction and grooming for provisioning of time-varying traffic in multilayer networks. In *27th International Conference on Optical Network Design and Modeling (ONDM)*. 2023.

[8] Wolpert, D. H. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.