Numerical and Statistical Probability Distribution Transformation for Modeling Traffic in Optical Networks

Alicja Poturała[1], Maria Konieczka[1],
Piotr Śliwka[2][0000−0001−6226−9580],
Sławomir Sujecki[3,4][0000−0003−4588−6741], and
Stanisław Kozdrowski[1,5][0000−0001−6647−5189]

[1] Department of Computer Science,
Faculty of Electronics, Warsaw University of Technology,
Nowowiejska 15/19, 00-665 Warsaw, Poland
[2] Department of Computer Science, Cardinal Wyszynski University
Woycickiego 1/3, 01-938, Warsaw, Poland.
[3] Telecommunications and Teleinformatics Department,
Wroclaw University of Science and Technology,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland.
[4] Faculty of Electronics, Military University of Technology,
S. Kaliskiego 2, 00-908 Warsaw, Poland.
[5] `s.kozdrowski@elka.pw.edu.pl`

Abstract. It is important for optical network operators to consider the available budget in forecasting network traffic. This is related to network expansion and equipment purchases. The underlying motivation is the constant increase in the demand for network traffic due to the development of new access technologies (5G, FTTH), which require particularly large amounts of bandwidth. The aim of this paper is to numerically calculate a transformation that allows determining probability distributions of network edge traffic based on known probability distributions of demand matrix elements. Statistical methods confirmed the proposed transformation. The study is performed for a practically relevant network within selected scenarios determined by realistic traffic demand sets.

Keywords: Optical Network Modeling and Optimization · Non-parametric and Parametric Probability Distribution · Statistical Analysis · Network Congestion · Forecasting · Traffic Demands.

1 Introduction

Optical network operators need to continually upgrade the networks to accommodate the ever-increasing data traffic. In currently deployed optical networks, which are based on single core fibres, the data transmission rate can be increased by using either a larger per-channel bit rate or by increasing the number of available channels [8]. In order to implement such changes in an operating optical

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

network a network operator needs to add equipment to the network nodes. This additional equipment incurs significant costs, which have to be accounted for in the company budget. It is needless to say that the minimisation of the equipment costs is critical to the commercial success of an operation network operator company. The cost minimisation is usually achieved by careful planning and optimisation of network resources at every stage of the network development. Algorithms for network optimisation have been developed for many years now [1, 3, 4, 9]. Of specific importance to optical networks are Routing and Wavelength Assignment (RWA) and Routing and Spectrum Allocation (RSA) algorithms applied in static [2, 7, 11] and dynamic [16] environment that has been developed by a number of authors over the past years. More recently, the predictive capability of a network optimisation software has been improved by taking into account the physical phenomena occurring in the optical fibre [8] and developing algorithms that predict the traffic demands in optical network nodes and corresponding data transmission rates in the network edges [10].

The traffic demands at optical network nodes increase steadily due to a continued modernisation of telecommunications technologies used in access networks. A milestone in the development of modern access networks was the introduction of glass fibres, which have now practically almost completely displaced in many countries the previously used copper connections. Fibre optics dominate the access network market in many countries, as they are superior to copper cable solutions by allowing larger data transmitting rates and longer distances. The other technology, which will further increase the traffic in access networks is 5G wireless technology, supported by its backhaul infrastructure. In the context of fast technological changes in access networks the analysis of demands and more specifically an estimation of the demand matrix elements is an essential element of the optical network planning and development.

In order to analyse and predict the values of demand matrix elements for optical network it is prudent to assume that a demand can be described by a random variable. Introducing a random variable, a probability distribution for a network matrix element can be empirically determined [10]. In the next step, one can attempt to fit parametric probability distributions to the empirically determined probability distributions [5]. In this context a natural question arises regarding the transformation of the probability distributions from the known probability distributions of demand matrix elements to the probability distributions describing the traffic in the network edges, which is the main problem considered in this contribution.

It is not easy to find an analytical form of the transformation from one nonparametric probability distribution of the random variable X to another nonparametric probability distribution of the random variable Y and expressed by the analytical function f such that $Y = f(X)$, where f is not a given explicitly. In some cases, for known parametric distributions, analytical transformations can be defined. However, to the best of the authors' knowledge, this problem has not found a satisfactory solution in the literature so far. Thus, one needs to

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

resort to the use of numerical methods, based on sampling or random number generators, and the methods of statistical data analysis.

The probability distribution transformation problem is therefore the primary objective of this contribution. It is noted that only after the traffic in the network edges is known the process of the optical network nodal equipment purchase planning can start. Also the use of statistical methods adds rigour to the way in which the problem is approached since it allows calculating formally expected values and the variance, i.e. the values and the ranges for the considered data transmission rate in a given network edge. We also provide a justification of the relationship between the known probability distributions of the demand matrix elements and the probability distributions describing traffic at the edges of the network.

The rest of the paper is organised as follows. In Section 2, the problem is described and the proposed methods for addressing the probability distribution transformation problem are presented. Next, in Section 3 the results obtained are presented together with the relevant discussion. Finally, Section 4, provides a summary of the main research findings.

2 Problem formulation and methods

A 3-node slice containing the nodes Wroclaw, Lodz and Katowice and the connections between them were selected from the Polish backbone network. The whole network together with the analysed part consisting of 3 nodes (marked using the red line) is shown on the Fig. 1.

In the following subsections the research methods are described. First, the approach used for the calculation of the demands matrix elements probability distributions is discussed. In the second subsection, the technique applied to calculate the probability distributions of the network edge data transmission rates subject to the known values of the demand matrix elements probability distributions is presented in a case of a 3 node network. In the last subsection an approach to calculating the probability distributions of the network edge data transmission rates for larger networks is discussed.

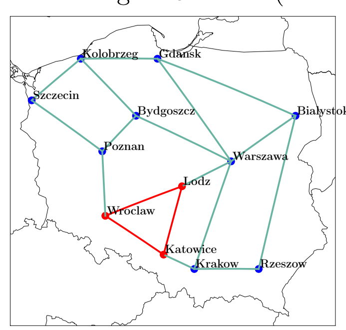


Fig. 1: Case of study, the segment of Polish core optical network.

2.1 Demands matrix elements probability distributions

When attempting to forecast the values of demands matrix elements, i.e., calculate their probability distributions in the coming years one needs historical data to build suitable stochastic models. However, such data are not generally available, as telecommunications network operators protect such information.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

Also SNDlib database lacks data on the historical evolution of demand matrix elements for the Polish network considered here [12]. Consequently, the values of the demands matrix elements had to be calculated using other methods. The general description of the method adopted here is described in [10]. The approach presented in [10] relies upon combining data from two statistical offices: Central Statistical Office (CSO) and European Statistical Office (Eurostat). Using the statistical data stemming from both sources the historical values of traffic demands for specific network nodes were estimated. Then the demands between pairs of cities of the Polish backbone network were estimated [10].

The Table 1 shows the calculated historical values of the demands matrix elements in the years 2010–2020 between cities of the considered network section consisting of 3 nodes (expressed in units of Tera bit per second [Tbps]). The demand is modeled by the sum of the values resulting from the trend present in the historical data ($\hat{y}_t$) and a random variable ($e_t = y_t - \hat{y}_t$) that represents the residuals, and is based on the relationship: $y_t = \hat{y}_t + e_t$ [15].

Table 1: Demands.

Year	Lodz–Kat.	Wro.–Kat.	Wro.–Lodz
2010	0,0909	0,0833	0,1211
2011	0,0972	0,0919	0,1303
2012	0,1099	0,0961	0,1389
2013	0,1100	0,0959	0,1393
2014	0,1125	0,1036	0,1501
2015	0,1099	0,1036	0,1496
2016	0,1184	0,1102	0,1576
2017	0,1191	0,1124	0,1600
2018	0,1202	0,1097	0,1580
2019	0,1228	0,1215	0,1684
2020	0,1319	0,1297	0,1801

Three methods were selected, based on the following procedure:

S1. on the basis of empirical data Y, estimate the parameters $\hat{\alpha}$ of the appropriate trend function $\hat{Y} = f(\hat{\alpha}, t)$, determine residuals $e = \hat{Y} - Y$ and finally the probability distribution function of the value of random variable Y as one realisation of an appropriate stochastic process,
S2. generate a sequence of $\{z_1, \ldots, z_n\} \in Z$ random numbers based on $F^{-1}(Y) := Z$, where F - distribution function of the random variable Y,
S3. based on a sequence of $\{z_1, \ldots, z_n\}$ from S2. find the "averaged" probability distribution of the random variable Z,

These methods calculate the probability distributions of the demands matrix elements using:

1. Extended Empirical Distribution (EED) - a stationary distribution that is computed from data containing historical residuals (e_t) and the observations generated from them. Limited historical data are available, so the idea is to estimate 1000000 new ones using the inverse transform sampling. All points are divided into k bins. Each bin represents one class with probability proportional to the number of samples;
2. Normal Distribution (ND) - in contrast to EED, describes the variables using μ and σ parameters (computed from e). For discretisation, one million points were generated from the $N(\mu, \sigma)$ distribution, which were divided into k bins;
3. Model with Increasing Uncertainty (MwIU) - took into account that predictions become less and less reliable with the length of the prediction horizon,

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

which is used in fan-based methods ([5],[14]). They are typically based on the two-part normal distribution [6] represented in the equation (1), where $A = \frac{2}{(1/\sqrt{1-\gamma})+(1/\sqrt{1+\gamma})}$.

$$f(x;\mu,\sigma,\gamma) = \frac{A}{\sqrt{2\pi}\sigma} \begin{cases} exp\{-\frac{1-\gamma}{2\sigma^2}[(x-\mu)^2]\}, & \text{for} \quad x \leqslant \mu \\ exp\{-\frac{1+\gamma}{2\sigma^2}[(x-\mu)^2]\}, & \text{for} \quad x > \mu \end{cases} \tag{1}$$

This distribution has three parameters mode - μ, uncertainty indicator - σ and inverse skewness indicator - γ. Nonstationarity was introduced by multiplying the uncertainty coefficient values in successive years. The three-sigma rule of thumb was used for discretisation.

2.2 Calculation of edge data transmission rate probability distributions for a small network

A demand matrix is used to describe the data transmission demands across the DWDM network. Its example form for a 3-node network is shown in the equation (2). This matrix is a square matrix of dimension N x N, where N denotes the number of nodes in the network. The elements from row l and column m contains the data transfer demand between city l and m expressed in Gbps.

$$D^{ijk} = \begin{bmatrix} 0 & d^i_{12} & d^j_{13} \\ 0 & 0 & d^k_{23} \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

The demand matrix is a multivariate random variable. In the approach adopted here, it is assumed that each of its elements is ultimately described by a discrete probability distribution consisting of c classes (or bins), where $c \in \{3, 4, 5, 6\}$. In the equation (2), the indices i, j, and k are used to specify the class to which each element of the demand matrix belongs, where $i, j, k \in \{0, \cdots, c-1\}$. The label 0 corresponds to the minimum class and $c-1$ to the maximum class. To fully describe a random variable next to the values (i.e. data transfer demand between two selected cities expressed in Gbps) one needs to specify the corresponding values of the probability that the random variable assumes a specific value. The individual elements of the matrix D^{ijk} have a corresponding probability of occurrence. Thus, a matrix of the probabilities for occurrence of a specific value of the matrix D^{ijk} element are given by the corresponding elements of the matrix P^{ijk}:

$$P^{ijk} = \begin{bmatrix} 0 & p^i_{12} & p^j_{13} \\ 0 & 0 & p^k_{23} \\ 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Thus, knowing the values of the matrix P^{ijk} elements and assuming that the random variables corresponding to the elements of matrix D^{ijk} are independent of each other one can calculate the probability of a specific matrix D^{ijk} realisation as the product of the corresponding matrix P^{ijk} elements. Once all the demands matrix elements are set and the corresponding probability of the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

specific realisation of demands matrix D^{ijk}, is known one can start calculating the values of the data traffic in the network edges by solving the optimisation problem as described in [8] subject to the known constrains. This procedure has to be repeated for each possible realisation of the demands matrix D^{ijk} to give the full probability distribution of the data traffic in the network edges, which in the considered 3 node example, and including 4 classes for matrix D^{ijk} elements discrete probability distributions gives 4^3 optimisations to calculate the full probability distribution of network edge data transmission rates.

2.3 Extending edge calculations to complex networks

The advantage of studying a small network segment is that it allows for a detailed estimation of network traffic based on demands matrix element forecasts. However, the accuracy of such approach is limited since other nodes of the network have an impact on the traffic present in the specific segment. Unfortunately, a full analysis of all realisations of the demand matrix in case of an entire backbone network using the presented approach leads to very intensive computations. First of all, it should be noted that the number of combinations of demand matrices is $c^{\frac{N(N-1)}{2}}$, where c is the number of classes used in probability distributions while N is the number of network nodes. So, even with a 12-node network and a three-class distribution, the number of combinations of demand matrices is 3^{66}. Moreover, as the complexity of the network rises, the time required to find a solution to the optimisation problem increases. For these reasons, an alternative approach has been considered giving an additional insight into the network operation. It consists in analysing the impact of changes in demands for selected city pairs. For this purpose the demand matrix element corresponding to the considered city pair is assumed to be described by a random variable whilst all other demands matrix elements are approximated by the expected value of the corresponding random variable. As an illustrative example in the next section an analysis was carried out of data traffic between selected distant city pairs: Szczecin-Rzeszow and Rzeszow-Poznan.

3 Results

The first series of experiments started by generating demand matrices for all class numbers and for all methods of forecast considered as described in section 2.1. The demand matrix element values corresponding to an example pair of cities (Katowice and Lodz) with a four-class distribution are collected in Table 2. The first two distributions (EED and ND) have similar expected values, which is due to the fact that they use the same linear model and the mean values of the deviations are close to zero.

In order to assess the quality of the predictions, the Theil Index values (I^2) and relative prediction errors *„ex post“* ($v_{s_f} = \frac{\sqrt{\frac{1}{\#I_f}\sum_{t\in I_f}(y_t - y_t^f)^2}}{\bar{y}_{t\in I_f}}$, where y_t^f - forecast of y in time t, I_f - forecast verification time interval) of the distributions

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

Table 2: Comparison of forecast demand levels between Katowice and Lodz, 4-class distribution, (probability).

Id	Year	Levels			
		Min	Negative	Positive	Max
EED	2019	616(0.278)	638(0.278)	653(0.222)	679(0.222)
	2020	633(0.278)	655(0.278)	670(0.222)	696(0.222)
	2021	650(0.278)	672(0.278)	687(0.222)	713(0.222)
	2022	667(0.278)	689(0.278)	704(0.222)	730(0.222)
	2023	684(0.278)	706(0.278)	721(0.222)	747(0.222)
ND	2019	556(0.010)	623(0.491)	669(0.494)	743(0.005)
	2020	573(0.010)	640(0.491)	686(0.494)	760(0.005)
	2021	590(0.010)	656(0.491)	703(0.494)	777(0.005)
	2022	607(0.010)	673(0.491)	720(0.494)	794(0.005)
	2023	624(0.010)	690(0.491)	737(0.494)	811(0.005)
MwIU	2019	587(0.065)	630(0.431)	659(0.431)	702(0.065)
	2020	576(0.065)	640(0.431)	683(0.431)	748(0.065)
	2021	564(0.065)	650(0.431)	707(0.431)	794(0.065)
	2022	552(0.065)	660(0.431)	732(0.431)	839(0.065)
	2023	540(0.065)	670(0.431)	756(0.431)	885(0.065)

employed were collected in Table 3. Since there were no significant differences between the expected values, the relative prediction errors and Theil Index values are similar for the distributions considered. The least accurate predictions were obtained for the Katowice-Wroclaw demand (although most values were acceptable, as $5\% < v(s_P) < 10\%$, where a commonly acceptable level of good fit is $v(s_P) < 10\%$ [15]). The forecasts' accuracy for the remaining demands is very good ($v(s_P) < 3\%$). Similar conclusions can be drawn by considering the Theil coefficient.

Table 3: Comparison of I^2 and $v(s_P)$ values of generated forecasts for the studied distributions for 3-node network.

Bin	Demand	I^2			$v(s_P)$		
		EED	ND	MwIU	EED	ND	MwIU
3	Katowice-Lodz	0.00053	0.00053	0.00054	0.02296	0.02306	0.02332
	Katowice-Wroclaw	0.00240	0.00285	0.00286	0.04905	0.05340	0.05351
	Lodz-Wroclaw	0.00066	0.00072	0.00072	0.02563	0.02680	0.02681
4	Katowice-Lodz	0.00054	0.00056	0.00054	0.02322	0.02376	0.02324
	Katowice-Wroclaw	0.00275	0.00282	0.00282	0.05252	0.05312	0.05316
	Lodz-Wroclaw	0.00083	0.00073	0.00071	0.02877	0.02710	0.02675
5	Katowice-Lodz	0.00054	0.00054	0.00055	0.02329	0.02327	0.02357
	Katowice-Wroclaw	0.00275	0.00285	0.00285	0.05251	0.05338	0.05339
	Lodz-Wroclaw	0.00082	0.00071	0.00071	0.02865	0.02675	0.02672
6	Katowice-Lodz	0.00055	0.00055	0.00054	0.02338	0.02349	0.02332
	Katowice-Wroclaw	0.00274	0.00286	0.00286	0.05235	0.05355	0.05351
	Lodz-Wroclaw	0.00084	0.00073	0.00073	0.02904	0.02713	0.02697

The Kolmogorov-Smirnov goodness of fit test was used to verify the statistical concordance of the probability distributions: demand matrix elements and traffic in the network edges for the 3rd to sixth classes. Tables 4–5 contain the results of the KS statistics given by the formula 4.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

$$\sqrt{\frac{n \cdot n}{2n}} \cdot \sup_x |F_{ref}(x) - F_{id}(x)| \quad (4)$$

where

$F_{ref}(x)$ - the distribution function of the demand matrix elements,

$F_{id}(x)$ - the distribution function of the forecast demand level (given e.g. in table 2).

Depending on the number of classes, the empirical KS_{emp} statistics values obtained in Tables 4-5, with a significance level $\alpha = 0.05$, do not exceed $KS_{theor} = 1.358$. So there are no grounds for rejecting $H_0 : F_{ref}(x) = F_{id}(x)$ against the alternative $H_1 : F_{ref}(x) \neq F_{id}(x)$. At the significance level $\alpha = 0.05$ we can assume that the respective probability distributions are consistent. As expected, an increase in the bin number generally increases the value of KS_{emp}, which in the case of high granularity of the probability distribution may lead to the relationship $KS_{emp} > KS_{theor}$, i.e. rejection of H_0 (a lack of accordance between the compared distribution functions [13]).

Table 4: Values of KS tests for stationary distributions.

Id	Bin	Demand			
		Katowice-Lodz	Katowice-Wroclaw	Lodz-Wroclaw	Max per demand
EED	3	0.0001	0.0010	0.0005	0.0010
	4	0.0789	0.0012	0.1566	0.1566
	5	0.1052	0.0691	0.2804	0.2804
	6	0.0005	0.0017	0.0009	0.0017
ND	3	0.4696	0.6177	0.3356	0.6177
	4	0.4571	0.7071	0.4578	0.7071
	5	0.4624	0.6201	0.2789	0.6201
	6	0.6750	0.4887	0.6743	0.6750

Table 5: Values of KS tests for MwIU.

Year	Bin	Demand			
		Kat.-Lodz	Kat.-Wro.	Lodz-Wro.	Max per demand
2022	3	0.3512	0.4872	0.2151	0.4872
	4	0.7071	0.7071	0.7071	0.7071
	5	0.4327	0.4457	0.4327	0.4457
	6	0.6736	0.8660	0.6736	0.8660
2023	3	0.3512	0.4872	0.2151	0.4872
	4	0.7071	0.7071	0.7071	0.7071
	5	0.4327	0.4457	0.4327	0.4457
	6	0.6736	0.8660	0.8660	0.8660

As expected, when the bin number increases, the value of KS_{emp} increases. It means that the higher number of bins, the greater probability of rejecting H_0, (which is in line with the intuition) the greater accuracy of the probability distribution, expressed with more classes. This may, from a statistical point of view, lead to a lack of accordance between the compared distribution functions.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

Table 6 presents the values of the Kolmogorov-Smirnov statistics scaled by class count for the 2019 and 2023 MwIU distributions for different bins. The comparison of the remaining pairs of years had similar yields similar results and for this reason it is not included in the Table 6.

Table 6: Values of KS tests for MwIU between MwIU.

Years	Bin	Demand			
		Kat-Lodz	Kat-Wro	Lodz-Wro	Max per demand
2019 2023	3	0.8384	0.8384	0.8384	0.8384
	4	0.6143	0.6143	0.6143	0.6143
	5	0.7158	0.7158	0.7158	0.7158
	6	0.8289	0.8289	0.8289	0.8289

All generated realisations of the demand matrices served as input to the optimisation process as described in sections 2.2 and 2.3. In total, more than 10000 experiments were performed, which were based on the results of the forecasts from all distributions for 2019–2023. The simulations were designed to calculate the probability of congestion on the network and to identify the edges that are most heavily loaded. First we consider the 3 node network.

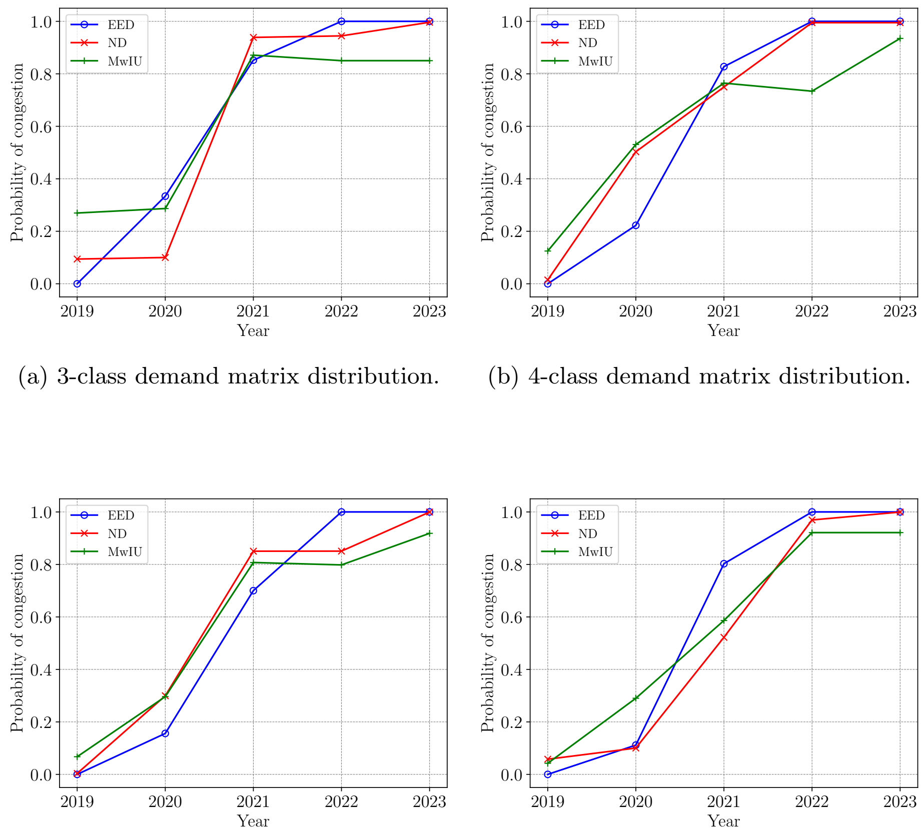
Fig. 2 shows the probability of network congestion for 3-(Fig. 2a), 4-(Fig. 2b), 5-(Fig. 2c) and 6-class (Fig. 2d) distribution, respectively. Each graph shows how the probability of network congestion has changed over the forecast years.

It can be seen that the probability of network congestion calculated using the empirical distribution (EED) ranges from 0 to 100%. This is because the differences between the lowest and highest predicted values are small - in 2019. For every realisation of the demand matrix, the algorithm found a solution to the problem in 2019 while in 2023, for every realisation of the matrix no acceptable solution could be found. Using the parametric distribution (MwIU) congestion probability never reached 100% while for the normal distribution (ND) and 2023, the congestion probability was very close to 100%.

For almost all distributions and class numbers, an increase in the congestion probability was observed in subsequent years. Only for the model with increasing uncertainty (MwIU) and the 3-, 4- and 5-class distributions, there was a decrease between 2021 and 2022. This is because in this case the range of accepted values widens within the prediction horizon.

As expected, the larger the number of classes into which the data was divided, the more accurately the probability of network congestion is estimated. As the number of classes increases, the ranges associated with individual classes become narrower. Also with the number of classes the number of possible realisations of the demand matrix increases exponentially. Thus, computational effort is needed to perform simulations. On the other hand, the shapes of the cumulative distribution function, regardless of the number of classes, stay almost unchanged.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32



(a) 3-class demand matrix distribution. (b) 4-class demand matrix distribution.

(c) 5-class demand matrix distribution. (d) 6-class demand matrix distribution.

Fig. 2: Probability of network congestion in subsequent years.

Based on the results obtained, it was decided that all three methods would also be selected for the application of the forecasts to the full network, as in the case of the 3-node network: EED, ND, and MwIU. In addition, the forecast years have been limited to 2019 and 2020. An in-depth analysis with the selected methods was carried out only for demands between distant city pairs: Szczecin-Rzeszow and Rzeszow-Poznan.

The Table 7 collects the obtained probabilities of realised classes for individual demands depending on the idea under study for 3-class distribution. The occurrence of underestimation is marked with a plus sign (+). As in the case of the forecast for the 3-node network, the EED predicted demand for 2020 is lower than in the case of ND and MwIU.

Tables 8a and 8b show a comparison of expected values calculated using selected probability distributions between years 2019 and 2023, for the demand matrix elements corresponding to Rzeszow–Szczecin and Szczecin–Poznan net-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

Table 7: Comparison of the probabilities of the class corresponding to the actual demand for analyzed city pairs [%].

Demand	Year	Distribution		
		EED	ND	MwIU
Rzeszow-Poznan	2019	22.21	90.33	15.69
	2020	+	5.66	15.69
Szczecin-Rzeszow	2019	22.21	88.75	15.69
	2020	+	5.53	15.69

work edge, respectively. As with the network slice analysis in Fig. 1, calculations were performed for all realisations of the demand matrix. The elements of the demand matrix were increased 15000 times to highlight the differences between the realisations under the adopted network parameters. Additionally, it is noted that the city pairs that were selected for in-depth analysis were characterised by small demand values.

Table 8: Comparison of expected values

(a) Rzeszow-Szczecin

Year	Distribution		
	EED	ND	MwIU
2019	1380.104	1380.373	1379.000
2020	1419.104	1419.373	1417.842
2021	1458.104	1458.373	1456.842
2022	1497.104	1497.333	1495.842
2023	1535.659	1535.430	1534.158

(b) Szczecin-Poznan

Year	Distribution		
	EED	ND	MwIU
2019	1139.433	1138.879	1139.000
2020	1175.433	1174.879	1174.842
2021	1210.765	1209.936	1210.000
2022	1246.765	1245.879	1246.000
2023	1282.433	1281.879	1282.000

Table 9 lists the I^2 and $v(s_P)$ values for the analyzed demands in the full Polish network for the 3-class distributions. Similarly to the studied network slice, the values were low enough to consider that the forecasts are either good or very good.

Table 9: Comparison of I^2 and $v(s_P)$ values of generated forecasts for the studied 3-class distributions for full Polish network.

Demand	I^2			$v(s_P)$		
	EED	ND	MwIU	EED	ND	MwIU
Rzeszow-Poznan	0.00093	0.00093	0.00096	0.03060	0.03053	0.03095
Szczecin-Rzeszow	0.00069	0.00069	0.00070	0.02627	0.02635	0.02641
sum	0.00162	0.00162	0.00165	0.05687	0.05688	0.05736

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

The representative results of numerical simulations for the year 2023 using demand matrix with 3-class probability distributions are shown in Fig. 3 and 4. The top indices of the demand matrix correspond respectively to the class indices for the Rzeszow-Poznan and Szczecin-Rzeszow pairs. Maps showing the bandwidth occupancy at each edge for the extreme values of demand matrix elements and the three considered approaches (EED, ND and MwIU) are presented in Fig. 4. Fig. 3 shows the channel occupancies calculated for the year 2023. The results obtained show that all distributions have similar demand expectation values.

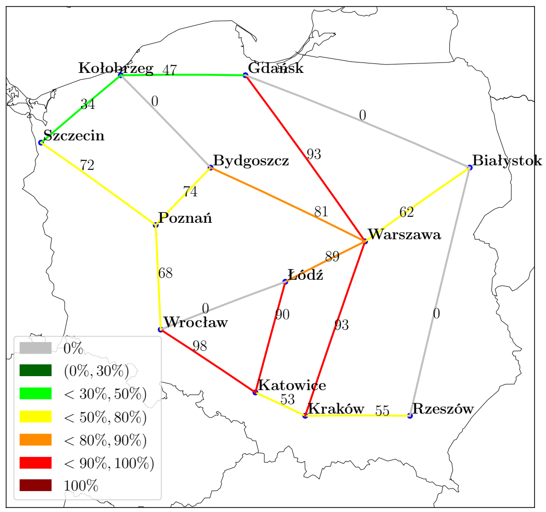


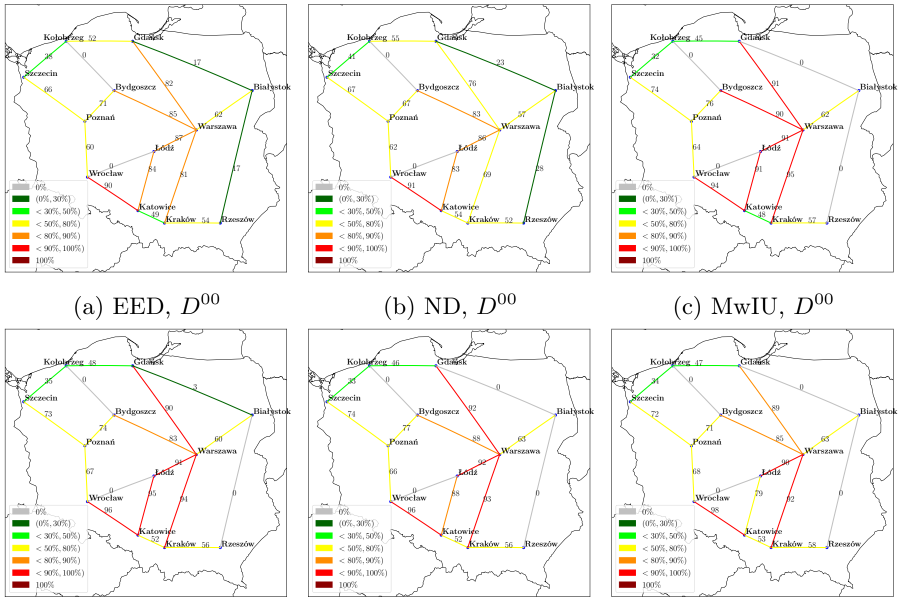
Fig. 3: Percentage edge occupancies for the 2023 forecast for D^{11} demand matrix, EED, ND and MwIU, 3-class distribution.

Also, it is noted that in all maps shown in Fig. 4 and 3 one can see that traffic tends to accumulate in the Warsaw node. Edges linked with Warsaw have the highest occupancy and the degree of this node is the only one equal to 5 while the degrees of all other nodes are lower. For the EED and ND distributions (Fig. 4a, 4d and 4b, 4e), it can be observed that the average channel occupancy per edge rises as the demands increase. In contrast, for the MwIU distribution this regularity is distorted (Fig. 4c, 4f). Although the demands are higher for MwIU D^{22} matrix , the average edge occupancy for this matrix was lower by about 0.6 pp than the result obtained for D^{00} matrix.

4 Conclusions

This contribution presents a statistical analysis of the data traffic in optical network and calculates estimates of the future values for the demand matrix

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32



(a) EED, D^{00} (b) ND, D^{00} (c) MwIU, D^{00}

(d) EED, D^{22} (e) ND, D^{22} (f) MwIU, D^{22}

Fig. 4: Percentage edge occupancies for the 2023 forecast depending on the demand matrix and the assumed 3-class distribution.

elements with parametric/non-parametric probability distribution. Once the demand matrix elements are known, an optimisation algorithm is used to predict the required network equipment needed to satisfy the traffic demand. As an example, firstly a 3 nodes network was considered and then full 12-nodes network with selected 2 demands. Depending on the nature of the data, three approaches to determining the types of distributions have been proposed: extended empirical distribution, normal distribution and model with increasing uncertainty.

The statistical concordance distribution test confirmed that the proposed numerical methods transform the empirical probability distribution into the probability distribution of demands and can be applied in the absence of analytical methods allowing for the transformation of the considered probability distributions. The challenge is still to provide the analytical function that allows for such a transformation of the probability distributions and a step further: transforming the empirical probability distributions into the transponder distribution.

References

1. Kalesnikau, I., Pióro, M., Poss, M., Nace, D., Tomaszewski, A.: A robust optimization model for affine/quadratic flow thinning: A traffic protection mechanism for networks with variable link capacity. Networks **75**(4), 420–437 (2020). https://doi.org/10.1002/net.21929

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32

2. Klinkowski, M., Żotkiewicz, M., Walkowiak, K., Pióro, M., Ruiz, M., Velasco, L.: Solving large instances of the RSA problem in flexgrid elastic optical networks. IEEE/OSA Journal of Optical Communications and Networking **8**(5), 320–330 (May 2016). https://doi.org/10.1364/JOCN.8.000320
3. Klinkowski, M.: Optimization of latency-aware flow allocation in ngfi networks. Computer Communications **161**, 344–359 (2020). https://doi.org/https://doi.org/10.1016/j.comcom.2020.07.044
4. Klinkowski, M., Walkowiak, K.: An efficient optimization framework for solving rssa problems in spectrally and spatially flexible optical networks. IEEE/ACM Transactions on Networking **27**(4), 1474–1486 (2019). https://doi.org/10.1109/TNET.2019.2922761
5. Konieczka, M., Poturała, A., Śliwka, P., Sujecki, S., Kozdrowski, S.: Modeling demands forecasts with probability distributions in dwdm optical networks. In: 2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). pp. 1–6 (2021). https://doi.org/10.23919/SoftCOM52868.2021.9559126
6. Kotz, S., Balakrishnan, N., Johnson, N.L.: Continuous multivariate distributions, Volume 1: Models and applications. John Wiley & Sons (2004)
7. Kozdrowski, S., Żotkiewicz, M., Sujecki, S.: Optimization of optical networks based on CDC-ROADM technology. Applied. Sciences. **9**(3) (2019). https://doi.org/10.3390/app9030399, http://www.mdpi.com/2076-3417/9/3/399
8. Kozdrowski, S., Żotkiewicz, M., Sujecki, S.: Ultra-wideband wdm optical network optimization. Photonics **7**(1) (2020). https://doi.org/10.3390/photonics7010016, https://doi.org/10.3390/photonics7010016
9. Kozdrowski, S., Banaszek, M., Jedrzejczak, B., Żotkiewicz, M., Kopertowski, Z.: Application of the ant colony algorithm for routing in next generation programmable networks. In: Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (eds.) Computational Science – ICCS 2021. pp. 526–539. Springer International Publishing, Cham (2021)
10. Kozdrowski, S., Sliwka, P., Sujecki, S.: Modeling traffic forecasts with probability in dwdm optical networks Springer Nature Switzerland AG, M. Paszynski et al. (Eds.): ICCS 2021, LNCS 12745, pp. 1–14, 2021. https://doi.org/10.1007/978-3-030-77970-2-28
11. Kozdrowski, S., Żotkiewicz, M., Wnuk, K., Sikorski, A., Sujecki, S.: A comparative evaluation of nature inspired algorithms for telecommunication network design. Applied Sciences **10**(19) (2020). https://doi.org/10.3390/app10196840, https://www.mdpi.com/2076-3417/10/19/6840
12. Orlowski, S., Wessäly, R., Pióro, M., Tomaszewski, A.: Sndlib 1.0-survivable network design library. Networks **55**(3), 276–286 (2010). https://doi.org/10.1002/net.20371
13. Sliwka, P.: Markov (set) chains application to predict mortality rates using extended milevsky–promislov generalized mortality models. J Appl Stat (2021). https://doi.org/10.1080/02664763.2021.1967891
14. Sliwka, P., Socha, L.: A proposition of generalized stochastic milevsky–promislov mortality models. Scand Actuar J **8**, 706–726 (2018). https://doi.org/10.1080/03461238.2018.1431805
15. Sliwka, P., Swistowska, A.: Economic forecasting methods with the R package. UKSW, Warszawa (2019)
16. de Sousa, A., Monteiro, P., Lopes, C.B.: Lightpath admission control and rerouting in dynamic flex-grid optical transport networks. Networks **69**(1), 151–163 (2017). https://doi.org/10.1002/net.21715

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_32