

Multivariate Machine Learning Models for Short-Term Forecast of Lightpath Performance

Stéphanie Allogba, Sandra Aladin and Christine Tremblay

Abstract—Machine Learning (ML) is emerging as a promising solution for managing the physical layer complexity of heterogeneous dynamic optical networks transporting multiple applications in a software defined network (SDN) context, namely for performance prediction. We propose two multivariate neural network models based on Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) methods, trained with field performance data and features, for predicting lightpath signal-to-noise ratio (SNR) over forecast horizons of up to 4 days. The best performance is achieved by using a 5-feature LSTM multivariate model over forecast horizons of up to 96 hours, with an absolute maximum error (AME) of 0.90 dB, compared to 0.91 dB and 0.97 dB for the GRU and LSTM univariate models, respectively, and 1.21 dB for a persistence model. The 2-feature multivariate models obtained through feature engineering perform better than their univariate counterparts for forecast horizons of up to 40 hours. Lastly, we explore the concept of transfer learning (TL) by testing the trained multivariate LSTM and univariate GRU models on field data from two lightpaths carried on the same route. The TL models underperform the naive model for the lightpath carried in a different optical fiber. However, for the lightpath carried in the same optical fiber on a portion of the same route, the LSTM-based TL model outperforms the naive model with a difference of up to 0.11 dB at a 96-hour forecast horizon, compared to 0.30 dB for the lightpath in the source domain, while using 3 times less training data.

Index Terms— Gated Recurrent Unit, Long Short-Term Memory, Machine Learning, Multivariate Neural Network, Performance Prediction, Quality of Transmission, Transfer Learning.

I. INTRODUCTION

TO deal with the constant growth in traffic, telecom operators deploy optical WDM transmission systems with ever-increasing data rates, capacity, and flexibility. However, as these systems' capacity increases and as they carry a multitude of applications, the impact of performance degradations and network failures is also greater. A potential way to reduce the impact of performance degradations and failures at the component, link, and network levels is to implement proactive network control and management tools

Manuscript submitted on May 25, 2021.

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grant RGPIN-2019-03972. The authors thank Thomas Tam and Tao Zhang from CANARIE for the field performance data.

S. Allogba, S. Aladin and C. Tremblay are with the Network Technology Lab, Department of Electrical Engineering, École de technologie supérieure, Montréal, Qc, Canada H3C 1K3 (e-mail: christine.tremblay@etsmtl.ca).

by leveraging the field data collected by the performance monitors deployed in the network. These tools can be implemented using ML-based performance predictors in a software-defined network (SDN) context.

In this paper, we propose two new multivariate QoT prediction models, expanding on the univariate LSTM and GRU models presented in [1], in order to forecast the QoT of an established lightpath. The two multivariate forecast models proposed in this study are trained using a database, herein referred to as KB-1. It consists of the historical BER data presented in [1], the channel received optical power (P_{RX}), the differential group delay (DGD) for one lightpath (hereafter called lightpath-1) carried on a 1300-km route in the CANARIE network, and the outside temperature and period of the day as additional features. Moreover, an analysis of the aforementioned features is carried out to study their impact on the performance of the multivariate models. Finally, the potential of transfer learning is explored by testing the trained multivariate SNR forecast models for lightpath-1 using two different datasets from two other optical lightpaths labeled lightpath-2 and lightpath-3, respectively. Lightpath-2 is carried on a 600-km segment of the 1300-km route, in the same direction as lightpath-1. Lightpath-3 is deployed in the opposite direction on the same 1300-km route as lightpath-1. The objective here is to evaluate two possibilities, one is the forecasting of the SNR of lightpaths that come from domains that have different configurations (opposite direction and different fiber on the same route) and the other, lightpaths that have identical configurations (same direction and same fiber on a portion of the route) of the source domain.

The paper is organized as follows. The first section provides an overview of the context of the paper while the following sections present the process of implementing the SNR prediction model as shown in Fig. 1. Thus, Section II describes the short-term SNR forecast problem as well as an overview of previous works in this area. The first step of the SNR prediction process is data preprocessing and it will be presented in Section III. This section includes a description of the lightpaths and databases. The first database (KB-1) is used for the construction of the multivariate ML models for lightpath-1 and the two others (KB-2 and KB-3) are used to evaluate the models on lightpath-2 and lightpath-3 using the transfer learning approach. Section III describes the management of missing data instances in the time series, and presents a statistical analysis of the data.

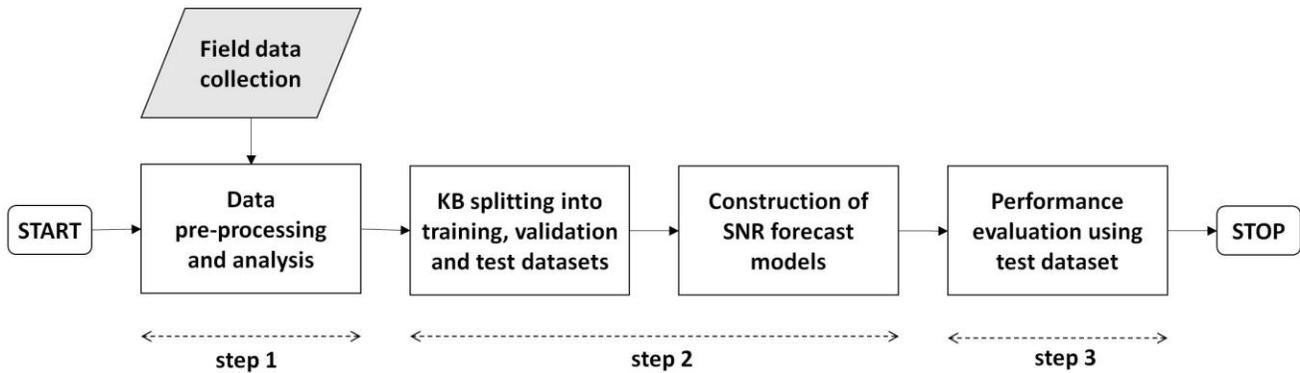


Fig. 1. SNR prediction methodology

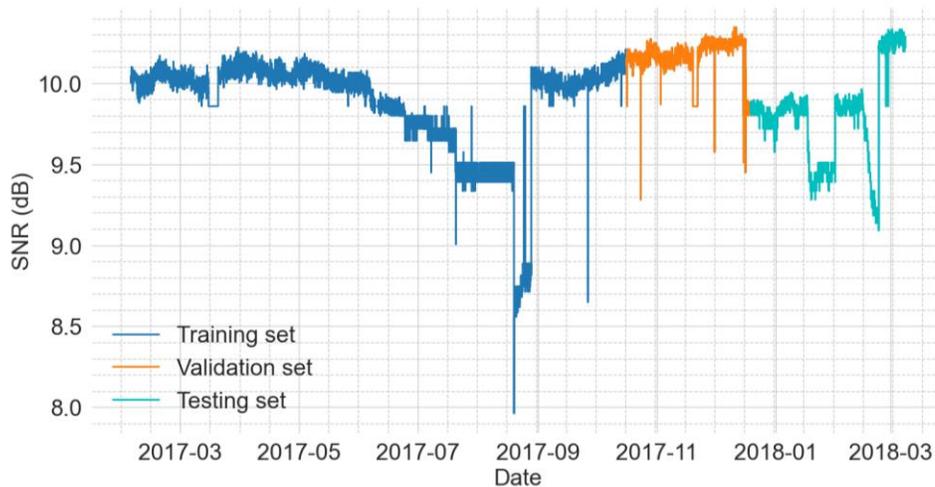


Fig. 2. Evolution of SNR for lightpath-1 in the CANARIE network over a 13-month observation period

The second step is the construction of the multivariate LSTM and GRU models (Section IV). This section includes a general description of the multivariate LSTM and GRU algorithms and hyperparameter optimization. The performance evaluation of the multivariate models trained with lightpath-1 data and the comparative analysis conducted with their univariate counterparts is also presented. Feature engineering is performed in Section V to assess the impact of each feature on the prediction accuracy. In Section VI, the transfer learning approach's validation is presented through a performance evaluation of the multivariate models constructed with lightpath-1 data on lightpath-2 and lightpath-3. Finally, the conclusions are drawn in Section VII.

II. SHORT-TERM SNR FORECAST PROBLEM

Today's optical networks are vast (measuring up to several thousands of kilometers) and heterogeneous (fiber and equipment types, outside plants, etc.), and are becoming more and more dynamic. Such network heterogeneity and complexity, together with incomplete topology and inventory databases, raise significant challenges, such as the effect of these combined factors on the performance of the lightpaths carried in these networks. Furthermore, the optical fiber itself is not only the best medium for transporting optical signals

over extremely long distances, it is also a good temperature and strain sensor. Lightpath performance can exhibit seasonal variations and degradations due to weather conditions, soil conditions (in the presence of ice, for example) and cable mishandling. Fiber cable aging can also cause loss increases over time. Accurate prediction of lightpath performance, either before or after their establishment, can be a complex task in such conditions as the accuracy of the analytical models used for such purposes is heavily dependent on the knowledge of system and link parameters and does not factor in external factors such as weather, outside plant characteristics and fiber/component aging.

Lightpath performance can be very stable over time. In such a case, short-term performance prediction is a very easy task that can be handled without the need for complex methods and tools. But some lightpaths exhibit a much more dynamic behavior that cannot be explained or predicted by theoretical models and switching effects. Fig. 2 shows a good example of a dynamic lightpath. On this figure, we can see the evolution of the SNR for lightpath-1 occurring in the CANARIE production network over a 13-month period. The SNR exhibits SNR degradations of up to several dB's during the spring and summer period, as well as several drops and increases in SNR during winter. These SNR variations last from a few days to

several months and cannot be explained by channel add-drop or switching.

Previous studies carried out on field data collected in aerial and buried links from different backbone networks in North America have revealed different patterns in the lightpath SNR time series: a seasonal behavior (i.e., lower SNR during the winter season); daily variations of polarization-dependent loss (PDL); and a much higher PDL activity during weekdays compared to weekends [2, 3].

The primary motivation of this work is to leverage machine learning (ML) to capture complex patterns in SNR time series and to exploit them for lightpath SNR forecast.

Machine learning (ML) has been studied for optical communication and network applications in the last few years [4, 5]. ML models have been explored for quality of transmission (QoT) estimation before lightpath establishment. Various supervised learning models trained with synthetic data have been proposed predicting the probability that the bit error rate (BER) will not exceed a predefined performance threshold [1, 6-8]. In [1], Artificial Neural Network (ANN) and Support Vector Machine (SVM) algorithms with different feature sets have been used to estimate the QoT of unestablished lightpaths. Random Forest (RF), linear and nonlinear regression models have been used to predict the actual BER value of a new lightpath to be established. This study, performed using a small knowledge base (KB) of 2,700 real BER samples, showed that RF tends to perform better than linear regression models [6]. ANN, RF, and logistic regression models have also been proposed to estimate the residual margin of a lightpath, with a better performance for ANN [7]. An NN-based algorithm has been proposed to estimate the lightpath optical signal-to-noise ratio (OSNR) for unestablished WDM channels using a synthetic KB of 1,400 samples [8].

ML methods can also be used for performance prediction. In the wireless domain, LSTM and GRU variants of deep recurrent neural networks (RNNs) have been proposed for link quality prediction [9]. In the optical domain, ML methods have also been used to predict lightpath performance using historical performance data [1, 10-12]. Two NN variants, namely GRU and LSTM algorithms, have been trained using monitored field data to predict signal-to-noise ratio (SNR) for an existing lightpath over forecast horizons of up to 24 hours [1, 10-12].

LSTM and GRU models provided better performance over very short horizons (8 hours and 12 hours) [11]. On the other hand, in [1, 10], LSTM and GRU models exhibited better performance for longer horizons (from 24 to 96 hours). More recently, a Convolutional Neural Network (CNN) was proposed to predict the performance of a lightpath using field data; the model was shown to capture the temporal SNR changes and to forecast it correctly over horizons of up to 24 hours [13]. The NN-based forecast models considered so far

were univariate, using single-lightpath historical field BER data to predict its future performance.

Moreover, the transfer learning approach has been used to predict or estimate the QoT of an unestablished lightpath [14-16]. In [14, 16], an ANN-based transfer learning model trained on a synthetic source domain for an optical system was used to predict the Q-factor on different optical systems, using datasets with similar distributions in the source and target domains. In [15], the authors used an SVM-based transfer learning model trained on synthetic data from a source domain to estimate the QoT of new optical lightpaths to be carried in a network with a different topology but using the same type of optical fiber and the same equipment.

To sum up, most of the ML applications at the physical layer have focused thus far on estimating lightpath QoT before establishment, using mainly synthetic data. Due to the scarcity of field data, less focus has been put on predicting the QoT of established lightpaths; the few models proposed so far are univariate LSTM and GRU models trained with historical BER data [1, 10, 12]. It should be noted that these univariate models only use historical SNR data from a single optical lightpath to predict SNR over a horizon of a few days. However, when no pattern is observable in the SNR time series, it becomes difficult for these ML models to predict the SNR data by taking into account only its temporal variations and to outperform a simple persistence model, as observed in previous work.

In this paper, our contributions are therefore organized as follows. First, we use multivariate deep RNNs models to take advantage of the available features in the field datasets for short-term SNR forecast. Then, based on the Pearson correlation test, we select the relevant features that can improve the performance and speed of the predictive models. Finally, we preliminarily experiment with transfer learning on two datasets of limited size. If successful, these predictive models could be used by network operators to trigger proactive responses to performance degradations.

III. KNOWLEDGE BASE

This section first presents a description of the databases, followed by the pre-processing phase and finally a statistical analysis of the time series data.

A. Databases description

The three databases (KB-1, KB-2 and KB-3) used in this study are built using the performance metric (PM) data from three different lightpaths deployed in a section of the CANARIE production network, as shown in Fig. 3.

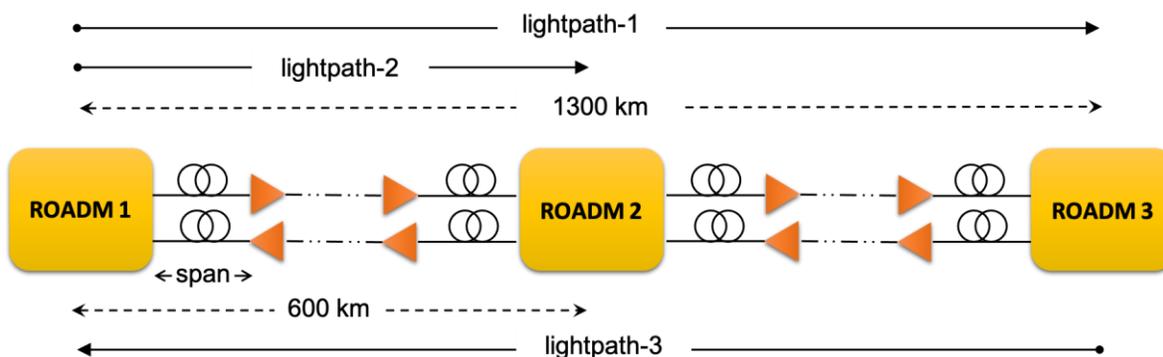


Fig. 3. Topology of the CANARIE network section (distances are approximate)

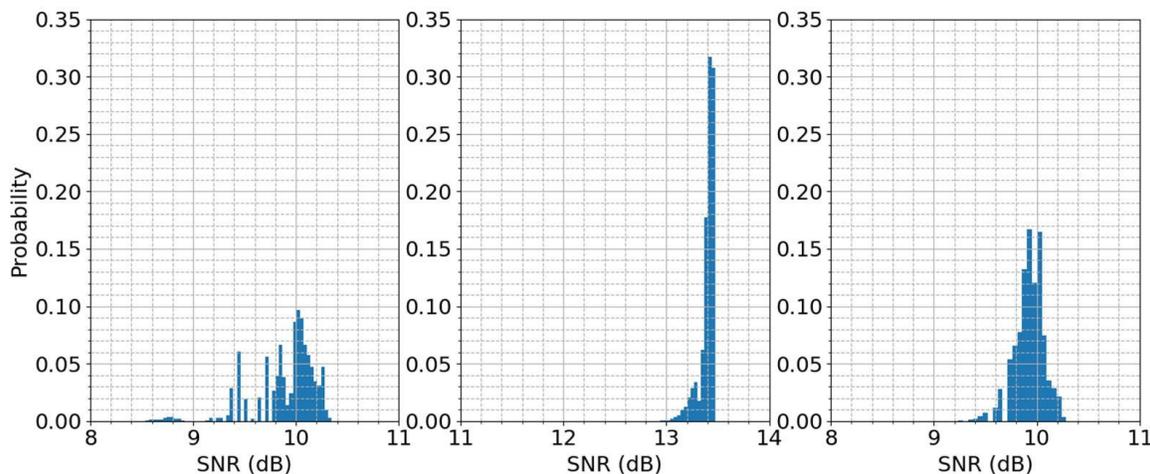


Fig. 4. Probability distribution function of SNR for: (a) lightpath-1; (b) lightpath-2; (c) lightpath-3

This network section comprises 3 reconfigurable optical add-drop multiplexer (ROADM) nodes interconnected with bidirectional optically amplified optical links. The average span length is 95 km, but no further details on the fiber, amplifier and fiber plant type are available. The network carries traffic using an unknown number of 100 Gb/s and 40 Gb/s channels. Lightpath-1 and lightpath-3 are two 100 Gb/s polarization multiplexed (PM) quadrature phase shift keying (QPSK) channels deployed on a bidirectional link of about 1300 km between ROADM 1 and ROADM 3. Lightpath-2 is a 40 Gb/s PM-QPSK channel deployed in a subsection of approximately 600 km between ROADM 1 and ROADM 2.

The collected PM data, monitored at a 15-minute sampling rate, include pre-forward error correction (pre-FEC) BER, P_{RX} and DGD for KB-1 and only the pre-FEC BER and P_{RX} for KB-2 and KB-3. The observation period is 13 months (i.e. from February 2017 to March 2018) for lightpath-1 [12], 5 months for lightpath-2 (i.e. from February 2017 to July 2017) and 18 months for lightpath-3 (i.e. from February 2017 to August 2018) [11]. Next, the SNR time series is computed from the raw pre-FEC BER data extracted from the PM data. The SNR time series includes 38,203 (KB-1), 16,000 (KB-2) and 17,832 (KB-3) data instances. For the purpose of this study, in order to consider KBs of similar size, the observation period for KB-3 was truncated from 18 months to 5 months

(i.e., from February 2017 to July 2017). The number of missing instances is 1,020 (i.e. 14.07% of KB) in KB-1, 704 (i.e. 4.42%) in KB-2 and 711 (i.e. 4.05%) in KB-3. The resulting SNR time series and probability distribution functions for the 3 lightpaths are shown in Fig. 4.

Missing values in time series can affect the performance of the prediction models [17]. Therefore, the missing data instances in the SNR, P_{RX} and DGD time series have been replaced with a moving average. Due to the sensitivity of information loss, as the time window of the moving average gets wider, it was set to 7.5 hours, as in [1]. In addition, note that when handling missing values, outliers, defined here as values exceeding 3 times the standard deviation, have not been removed from the time series.

In order to implement the multivariate models, two groups of features are considered as input to the models in addition to the historical SNR data. The first is from the monitored PMs, namely the power P_{RX} and DGD. The choice of this first group was made on the one hand because of the useful information that the receiver power can provide in the variation of the SNR, (indeed, the SNR is a function of the variation of the noise and of the signal) and on the other hand, we wished to be positioned in a real-world context in which the operator would have different monitored PMs. The addition of these extracted PMs could therefore make it possible to evaluate the relevance of adding monitored PMs to the models.

Table I. Lightpath and KB information

	KB-1	KB-2	KB-3
Lightpath	Lightpath-1	Lightpath-2	Lightpath-3
Link length	1300 km	600 km	1300 km
PM data	BER DGD P_{RX}	BER P_{RX}	BER P_{RX}
Features	SNR DGD P_{RX} Outside T Period of day	SNR P_{RX}	SNR P_{RX}
Number of instances*	38,203	16,000	17,832
Number of missing data instances	1,020	704	711
Observation period	13 months	5 months	18 months

*after filling the gaps in the time series

The second group consists of external factors related to the environment. Our objective is to predict gradual variations in the SNR, as observed in Fig. 2.

These variations could be caused by the physical properties of the network and could correspond to the deterioration or aging of the fibers caused either by external temperature acting on them, or by the central office's (CO) activities such as maintenance. Analyses that are mainly focused on data distribution and correlation studies, presented in [2, 18, 19], have indeed shown that there may be a relationship between BER and temperature as well as between BER and daily changes characterizing human activities in the CO. Thus, external factors are the hourly outside temperature (T) at the receiving site location and human activity. The temperature data ranged from -35 C and 34.4 C with a mean value of 2.6 C and a standard deviation of 14.5 C during the observation period [20]. As for human activity, it was defined by four periods of the day which represents the daily change: (5 am – 12 pm), (12 pm – 6 pm), (6 pm – 10 pm) and (10 pm – 5 am).

Table I summarizes the lightpath information and KB information also corresponding to the model input information.

B. Statistical analysis

The characteristics of the SNR, received optical power and DGD time series are shown in Table II. Note that the SNR variability for lightpath-2 and lightpath-3 is less than that for lightpath-1, with maximum SNR variation over the observation period of 0.08 dB and 0.39 dB, respectively, comparatively to 1.37 dB for lightpath-1.

Statistical analyses are done on the entire databases using the *statsmodels* package in Python 3. These analyses make it

Table II. Summary of statistical values of SNR, P_{RX} and DGD

	Lightpath-1	Lightpath-2	Lightpath-3
SNR (dB)			
min	7.96	11.12	9.04
max	10.34	13.47	10.28
average	9.90	13.39	9.91
sdev	0.29	0.07	0.17
P_{RX} (dBm)			
min	-6.69	-8.30	-7.5
max	-4.00	-6.19	-5.5
average	-5.45	-7.06	-6.15
sdev	0.49	0.33	0.37
DGD (ps)			
min	2	N/A	N/A
max	10	N/A	N/A
average	5.66	N/A	N/A
sdev	1.14	N/A	N/A

possible to better describe the data, which in turn makes it possible to improve the performance of the models.

The first analysis is the seasonality test. It identifies patterns in time series data as well as other components to see more clearly the relationship between model input (features) and output (data to be predicted). The time series is decomposed into trend, seasonal and residual components, using a Seasonal Trend decomposition based on Loess (STL). For example, STL tests were recently proposed to detect trends in optical span loss [21]. The decomposition was performed on the SNR time series using the *Statsmodel* library in Python 3. First, a peak-to-peak amplitude analysis revealed very little seasonality pattern for lightpath-1, with peak-to-peak amplitudes of 0.03 dB, and 0.08 dB for the 24-hour and 7-day seasonal components, respectively. Likewise, no significant seasonal component was found for lightpath-2 and lightpath-3 (0.06 dB for the 24-hour seasonal component). The strength of seasonality (F_s) was also calculated as in [22]. The F_s metric gives a measure of the strength of seasonality between 0 (no seasonality) and 1 (strong seasonality). F_s values of 0.003, 0.021, and 0.008 were found for lightpath-1, lightpath-2, and lightpath-3, respectively. This means that the daily and weekly seasonality information will not be useful to forecast the SNR.

The second analysis is the stationarity test, which consists of performing the Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test on the SNR time series. The purpose of the stationarity test is to determine whether the statistical properties of the time series data (mean, variance and covariance) do not change over time. A non-stationary time series is more difficult to predict because its statistical properties will probably not be the same in the future as in the past [22]. The results show that the SNR time series for lightpath-1 are not stationary (test values of -4.88 < critical values of -2.86 for ADF and 3.61 > critical value of 0.176 for KPSS). Similarly, the SNR time series for the two other lightpaths are not stationary, with test values of -4.25 <

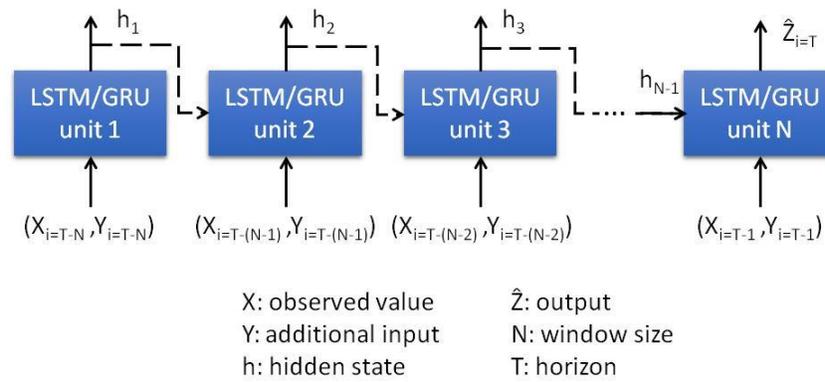


Fig. 5. Topology of the LSTM and GRU models

Table III. Optimal set of hyperparameters

	Multivariate model		Univariate model [1]	
	LSTM-M-1	GRU-M-1	LSTM	GRU
Number of units	512	512	256	256
Learning rate	0.00001	0.00005	0.00001	0.00001
Window size (h)	24	24	48	48
Dropout rate	0.2	0.2	0.2	0

critical value of -2.86 for ADF, and test values of 4.93 > critical value of 0.146 for KPSS (lightpath-2), and test values of -6.18 < critical value of -2.86 for ADF and test values 0.99 > critical value of 0.146 for KPSS (lightpath-3).

The advantage of using neural network algorithms and their variants is that these algorithms can predict the data without taking into account non-stationarity effects and without the need to use methods to make the time series stationary [10, 11, 23].

IV. MULTIVARIATE SNR FORECAST MODELS

This section is subdivided into two parts. The first presents the implementation of the multivariate LSTM and GRU models, referred to as LSTM-M-1 and GRU-M-1, respectively. Both models are built using KB-1. The second part presents the performance analysis of the models.

A. Architecture and parameter optimization

The multivariate LSTM and GRU models considered for short and long-term SNR forecasting are derived from neural networks. Multivariate models differ from their univariate counterparts by taking into account other features in addition to historical SNR data to predict future SNR values. To evaluate the impact of adding features on the accuracy of the

SNR prediction, our proposed multivariate models are compared to the univariate LSTM and GRU models in [1].

As described in [1], the univariate LSTM and GRU models use structures called gates. These gates control the cell states and the combination of inputs to determine the desired outputs. The desired outputs represent the SNR value to be predicted over the horizon T .

The topology of the LSTM and GRU models is illustrated in Fig. 5. As shown in Fig. 5, the multivariate LSTM and GRU models combine the observed values and other features as inputs, unlike the univariate models, which only use the observed value as input. In our proposed multivariate models, in order to predict the SNR of the lightpath at horizon T , the inputs are the channel received power, the DGD, the outside temperature, the period of day, as well as the SNR values over an observation period of N hours.

Similar to the univariate models in [1], the multivariate LSTM and GRU models are built using the *Keras* package in Python 3 and implemented by dividing the KB-1 into training, validation and test datasets according to the ratio of 0.64, 0.16, 0.20, respectively.

The training and validation datasets are used to train the models and to determine the appropriate hyperparameters, respectively. Also, using the same methodology as in [1], the hyperparameters such as the learning rate, dropout rate, and size of the hidden layer are optimized by testing over different values ((0.00001, 0.000025 and 0.00005), (0, 0.2, 0.5, 0.8 and 1), and (50, 150, 256 and 512), respectively) using the Root Mean Square Error (RMSE) as the performance metric.

In addition to the hyperparameters, the window size also has an impact on model performance. The window size corresponds to observation period N in the sequence of inputs used to forecast the next SNR at a given horizon T hours ahead. To determine the best window size, different values (1, 4, 8, 12, 24, 48 hours) are tested using the RMSE as the performance metric.

Table III shows the optimal set of resulting hyperparameters. For the multivariate LSTM-M-1 model, the learning rate is finally set at 0.00001, the size of the hidden layer at 512 neurons, the dropout rate at 0.2, and the window size at 24 hours. For the multivariate GRU-M-1 model, the learning rate is set at 0.00005, the size of the hidden layer at 512 neurons, the dropout rate at 0.2, and the window size at 24

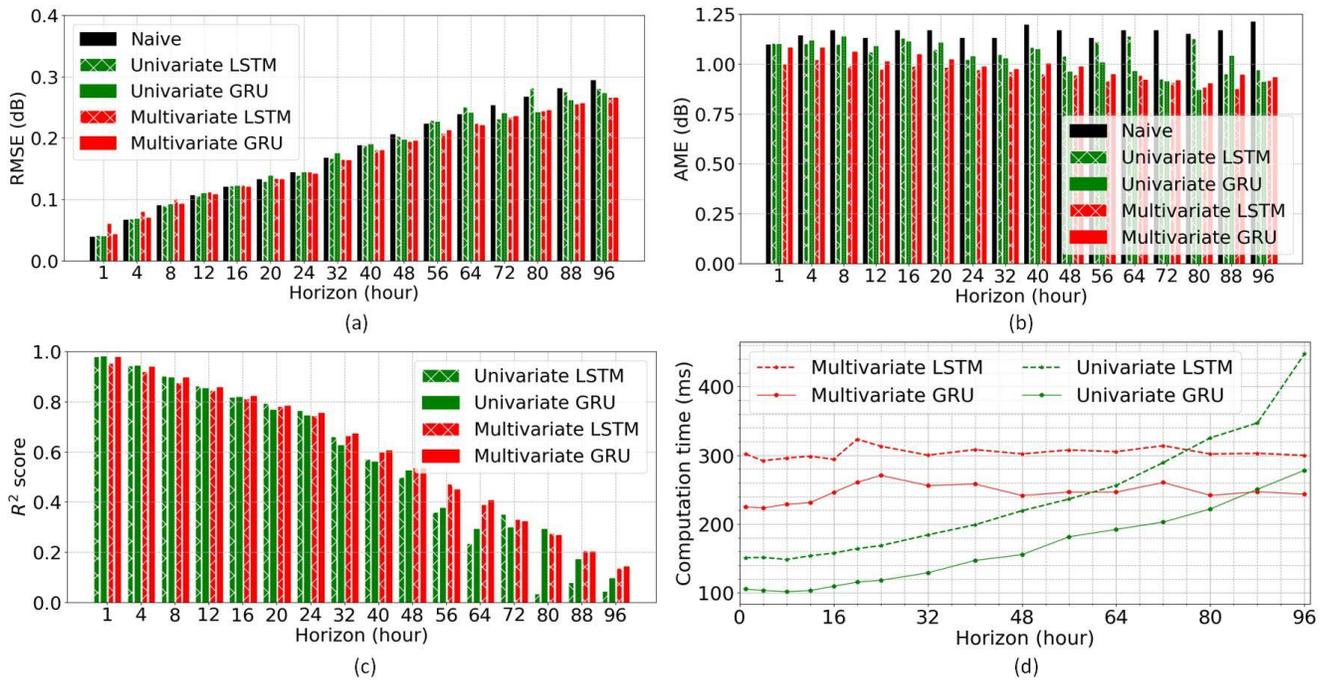


Fig. 6. Performance evaluation using the KB-1 test dataset: (a) RMSE; (b) AME; (c) R^2 score; (d) Computation time

hours. The univariate LSTM and GRU models used the same hyperparameters as in [1].

B. Performance results

After the models are built in the training step, they are evaluated using the test dataset of KB-1. The test dataset corresponds to a period of approximately 2.5 months (from December 2017 to March 2018). For the purpose of comparison, in addition to the LSTM and GRU models, a persistence (naive) model has also been implemented on the test dataset to evaluate the performance of the models as in previous works [24]. This naive method consists in assigning to the SNR value predicted at horizon T the latest value of the observation window N . Note that other baseline models such as linear regression can also be used as in [24]. However, since the main focus of this study was to compare univariate and multivariate models and to evaluate the impact of adding features on the prediction accuracy, other baseline models were not considered.

The four metrics used in the evaluation step are the RMSE to determine the accuracy of the models, the R^2 score to evaluate the robustness of the models, the absolute maximum error (AME) to evaluate the impact of the bad predictions on the model performance, and the computation time to compare the model execution times. Fig. 6 shows the performance results according to these four metrics for all the models. Note that the evaluation was performed for forecast horizons ranging from 1 to 96 hours.

The RMSE indicates how close the observed SNR values are to the predicted SNR values. Low RMSE values indicate better prediction. As shown in Fig. 6(a), the RMSE values increase with the forecast horizon for all the models, as expected. The RMSE of the multivariate and univariate

models are very similar for all forecast horizons combined, with a slight advantage for the multivariate models over the univariate and naive models as the forecast horizon increases. At 96-hour forecast horizon, the multivariate models performed slightly better than their univariate counterparts and outperformed the naive method by up to 0.03 dB.

The AME metric indicates the maximum error of the forecast model. Fig. 6(b) shows the AME for all the models. The lowest AME values were achieved with the multivariate models, with maximum prediction errors smaller than those of the univariate and naive models. Interestingly, the AME values for both multivariate models (and for the univariate GRU model) decrease as the forecast horizon increases, while the AME values for the naive model increase as the time horizon lengthens, showing the benefits of ML methods. The AME of the multivariate LSTM model ranged from 1.00 dB to 0.92 dB as the forecast horizon goes from 1 hour to 96 hours, compared to 1.10 dB to 1.21 dB for the naive model.

The R^2 score, presented in Fig. 6(c), determines the robustness of the models. It describes how well the selected features characterize the variability of the data to be predicted. Its values are less than or equal to 1, so that the closer the value is to 1, the more robust the model is. The R^2 score confirms the observations derived from the RMSE curves. The multivariate GRU-M-1 and LSTM-M-1 models performed very similarly and generally outperformed the univariate models. Note that for horizons of less than 24 hours, the R^2 score values for the multivariate and univariate models are very close. However, the higher the horizons, the more rapidly the univariate models become unstable, with an R^2 score that is less efficient than that of the multivariate models.

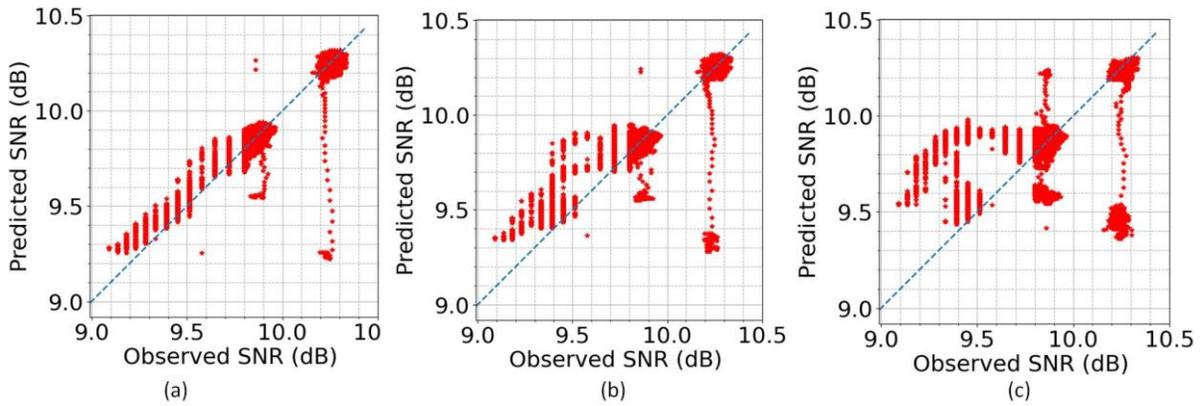


Fig. 7. Predicted SNR vs. observed SNR: (a) 1-hour forecast horizon; (b) 24-hour forecast horizon; (c) 96-hour forecast horizon

Table IV. Correlation analysis

	Channel SNR
Channel received power	0.80
Outside temperature	-0.20
DGD	-0.10
Period of the day	0.01

Table V. Optimal set of hyperparameters

	LSTM-M-2	GRU-M-2
Number of units	512	512
Dropout rate	0.2	0.2
Learning rate	0.00001	0.00005
Window size (h)	24	24

As shown in Fig. 6(d), the computation time is about the same for both multivariate models and does not increase much with the lengthening of the forecast horizon, while for univariate models it increases as the time horizon lengthens. Also, at longer forecast horizons, the multivariate models become faster than the univariate models. The models were run on a system with a 2.5 GHz Intel® Core™ i5-7200U processor, 8 GB RAM.

Fig. 7 shows the predicted SNR relative to the observed SNR at the shortest (1-hour) and at the longest (96-hour) forecast horizons as well as an intermediate (24-hour) horizon for the best performing model, namely the multivariate LSTM-M-1 model. The closer the predicted values are to the diagonal line, the more accurate the prediction is. Fig. 7 shows a good prediction accuracy at a short horizon and also, that the forecast accuracy decreases as the horizon increases. Note that the LSTM-M-1 model was unable to correctly predict the SNR values around 10.25 dB in the test dataset. This can be explained by the fact that such high SNR values in the test dataset fall outside the range of SNR values in the training dataset, as shown in Fig. 2. The scatter plots widen as the time horizon lengthens as a result of the reduction in prediction accuracy. The model also struggled to predict the relatively sharp SNR drops in the test dataset, which led to an overestimation of the predicted SNR values.

V. MULTIVARIATE SNR FORECAST MODELS WITH FEATURES REDUCTION

In this section, we study the impact of the different features on the performance of the multivariate models with the aim of reducing the number of features in the multivariate models.

A. Feature engineering and reducing the multivariate models

A correlation analysis using Pearson tests was first performed on the feature set. The purpose of the correlation analysis is to evaluate the importance of the different features in KB-1 in the hopes of reducing the feature set to only the most important ones to be used in the multivariate models.

Table IV presents the results of the Pearson tests. The tests reveal that the received channel received power is highly correlated to the SNR (correlation value = 0.8). A much smaller correlation is found with outside temperature, DGD and the period of the day.

Based on these results, two multivariate LSTM and GRU models (hereafter referred to as LSTM-M-2 and GRU-M-2, respectively) were built using the channel SNR and P_{RX} over an observation period of N hours as the only two features in the prediction process. Finally, using the test dataset, the same performance metrics (namely the RMSE, R^2 score and AME) are used to compare the two models LSTM-M-2 and GRU-M-2 to the best performing 5-feature model (LSTM-M-1), the best performing univariate model (GRU) and the persistence model.

As in section IV, the models LSTM-M-2 and GRU-M-2 are built using the *Keras* package in Python 3 and the reduced feature set. The resulting KB is divided into training, validation and test datasets according to the same ratio (0.64, 0.16 and 0.20, respectively). Table V presents the hyperparameters found from the validation dataset.

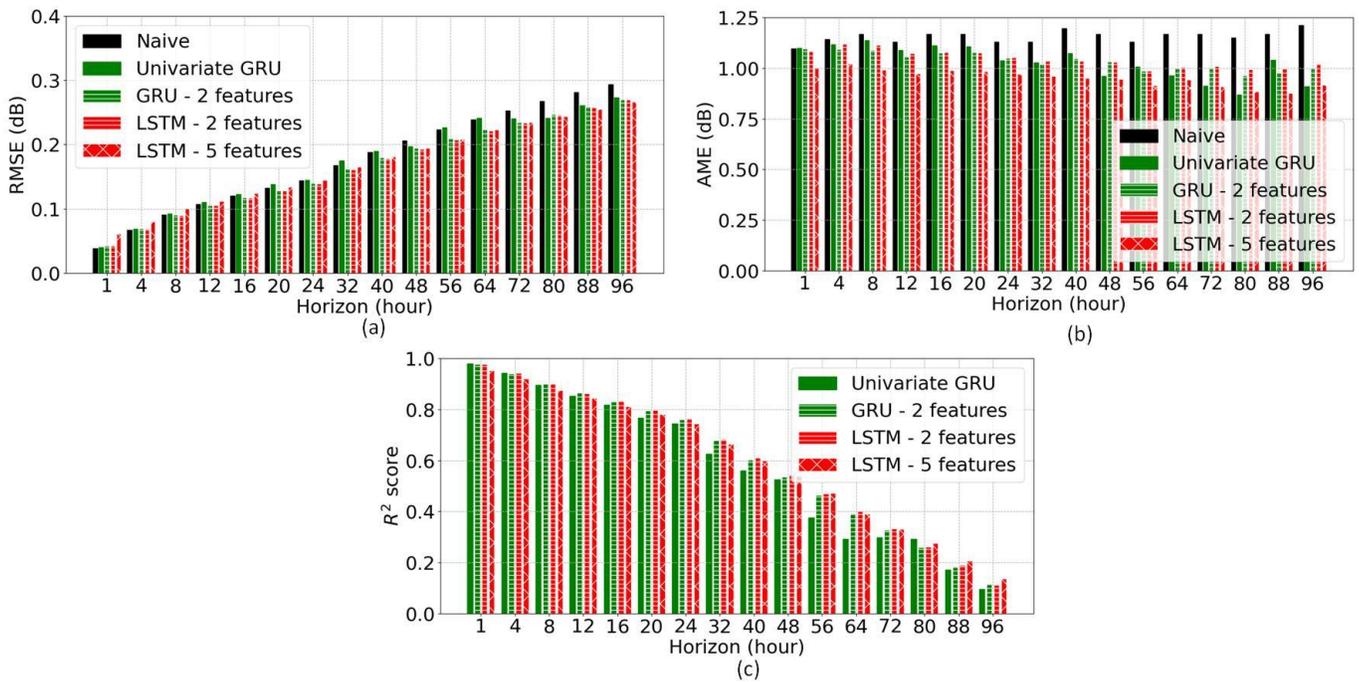


Fig. 8. Performance evaluation using a reduced feature set in KB-1: (a) RMSE; (b) AME; (c) R² score

Table VI. Comparative analysis of the univariate and multivariate models

Metric	Forecast horizon (hours)	Baseline	Univariate model [1]		Multivariate model			
		Naïve model	LSTM	GRU	LSTM-M-1	GRU-M-1	LSTM-M-2	GRU-M-2
		Features						
		SNR	SNR	SNR	SNR, DGD, P _{RX} , Outside T, Period of the day	SNR, DGD, P _{RX} , Outside T, Period of the day	SNR, P _{RX}	SNR, P _{RX}
RMSE (dB)	1	0.04	0.04	0.04	0.06	0.04	0.04	0.04
	24	0.14	0.14	0.14	0.14	0.14	0.14	0.14
	96	0.29	0.28	0.27	0.27	0.27	0.27	0.27
AME (dB)	1	1.10	1.11	1.10	1.00	1.08	1.08	1.10
	24	1.13	1.02	1.04	0.97	0.99	1.05	1.05
	96	1.21	0.97	0.91	0.92	0.93	1.02	1.00

Indeed, the hyperparameters, such as the learning rate, dropout rate, size of hidden layer and window size, are optimized by testing with the RMSE on the same values ((0, 0.00001, 0.000025 and 0.00005), (0, 0.20.5, 0.8 and 1), (50, 150, 256 and 512) and (1, 4, 8, 12, 24, 48 hours), respectively).

As a result, the learning rate is set at 0.00005, the size of the hidden layer at 512 neurons, the dropout rate at 0.2, and the window size at 1 hour for the multivariate LSTM model with 2 features, whereas for the multivariate GRU model with 2

features, the learning rate is set at 0.000025, the size of the hidden layer at 256 neurons, the dropout rate at 0.2 and the window size at 8 hours.

B. Performance results

The performance results of the LSTM-M-2 and GRU-M-2 models using a reduced feature set in KB-1 are shown in Fig. 8. As shown on the RMSE bars in Fig. 8(a), the models

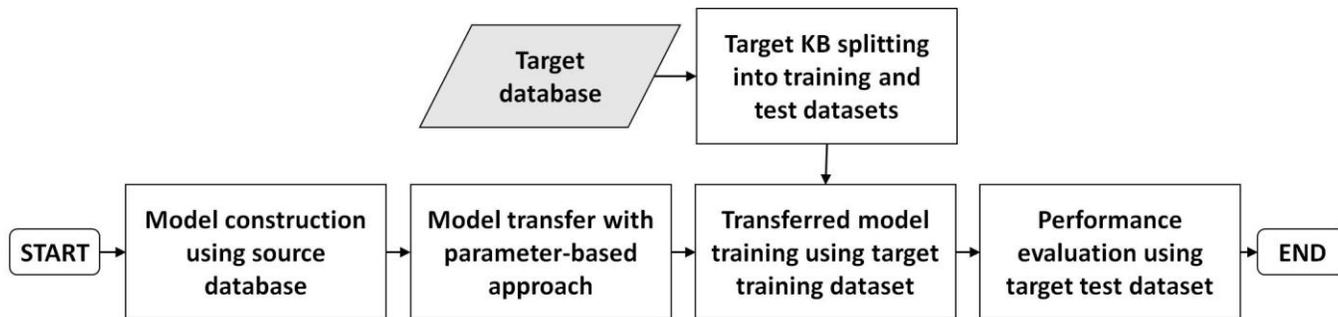


Fig. 9. Model fine-tuning process using transfer learning

LSTM-M-2 and GRU-M-2 performed similarly to their 5-feature counterparts LSTM-M-1. Both models outperformed the naive model with an RMSE difference of up to 0.02 dB.

Fig. 8(b) shows that the LSTM-M-1 model outperforms the 2-feature multivariate models LSTM-M-2 and GRU-M-2, with a lower (about 0.05 dB difference) AME over all time horizons combined. Moreover, the 2-feature models perform similarly and outperform the naive model.

Fig. 8(c) shows very similar R^2 scores for the 2-feature and 5-feature multivariate models.

The 2-feature multivariate GRU-M-2 and LSTM-M-2 models were executed on a system with a 2.5 GHz Intel® Core™ i5-7200U processor, 8 GB RAM and remained stable at approximately 15.5 ms and 16.1 ms, respectively. They are also faster than their 5-feature counterparts.

Table VI presents a comparative performance analysis of the variants of the multivariate models against univariate and naive models. The RMSE and AME values are shown for 1-hour, 24-hour, and 96-hour forecast horizons, with the best performance (lowest values) shown in bold characters. The first observation is that complex RNN models did not exhibit a RMSE benefit over a simple naive method at very short forecast horizons (24 hours or less). The slightly better RMSE performance of the univariate models over the naive model at 96-hour forecast horizon did not improve further by using multivariate models. However, according to the AME performance metric, the 5-feature multivariate models outperformed their univariate counterparts as well as the naive models across all forecast horizons, except at the 96-hour horizon where the univariate GRU model shows a 0.01-dB performance benefit. Furthermore, the higher the forecast horizon, the bigger the AME advantage over the univariate and naive models. The best overall performance across all forecast horizons, in terms of RMSE and AME, was obtained by using the 5-feature multivariate LSTM model. The multivariate GRU model should not be discarded as it exhibited a similar (although slightly lower) performance but a shorter calculation time. Finally, although the 2-feature multivariate models underperformed the 5-feature multivariate models and the univariate models, these models should not be discarded neither. The 2-feature models outperformed the naive method and could be run in much shorter execution time than the 5-feature multivariate models and univariate models, with up to a 95% reduction in computing time up in both cases.

VI. GENERALIZATION OF THE SNR FORECAST MODELS USING TRANSFER LEARNING

Transfer learning can be useful when the amount of data on hand to train the forecast models is limited. Thus, in this section, we experiment with the concept of transfer learning by testing the univariate and multivariate models under two scenarios using the KB-2 and KB-3 databases from the CANARIE network. The first scenario involves field data from a 40G PM-QPSK lightpath (lightpath-2) carried in the same optical fiber and deployed on a 600-km section of the same route [11].

The second scenario considers data from another 100G PM-QPSK channel (lightpath-3) carried in the opposite direction on the same 1300-km route as lightpath-1.

A. Description of the transfer learning approach

As shown in Fig. 9, the concept of transfer learning consists of transferring a model implemented in a given domain to another domain to answer a predefined problem. In the literature on transfer learning, a domain is defined by a feature space and a marginal probability distribution; furthermore, for a given domain, a task is defined by a label space and a predictive function [25]. The domain in which the model is implemented is called the source domain and the domain in which the problem needs to be solved is called the target domain.

By using transfer learning, our aim is to determine whether the forecast models trained with a specific lightpath (source domain) can be generalized. They could then be used to test channels (target domain) deployed in the same direction on a portion of the same optical fiber (scenario 1) and ultimately in the opposite direction on the same route, i.e., on a different optical fiber (scenario 2), under the assumption that the source and target domains are different.

In this work, the source domain dataset is KB-1 (lightpath-1) and the target domain datasets are KB-2 (lightpath-2) and KB-3 (lightpath-3). As shown in Fig. 3, lightpath-3 is carried between the same nodes as lightpath-1, but in the opposite direction. Lightpath-2 is carried in the same optical fiber on a shorter section (600 km) of the same route as lightpath-1. The target dataset distributions and variabilities are also different

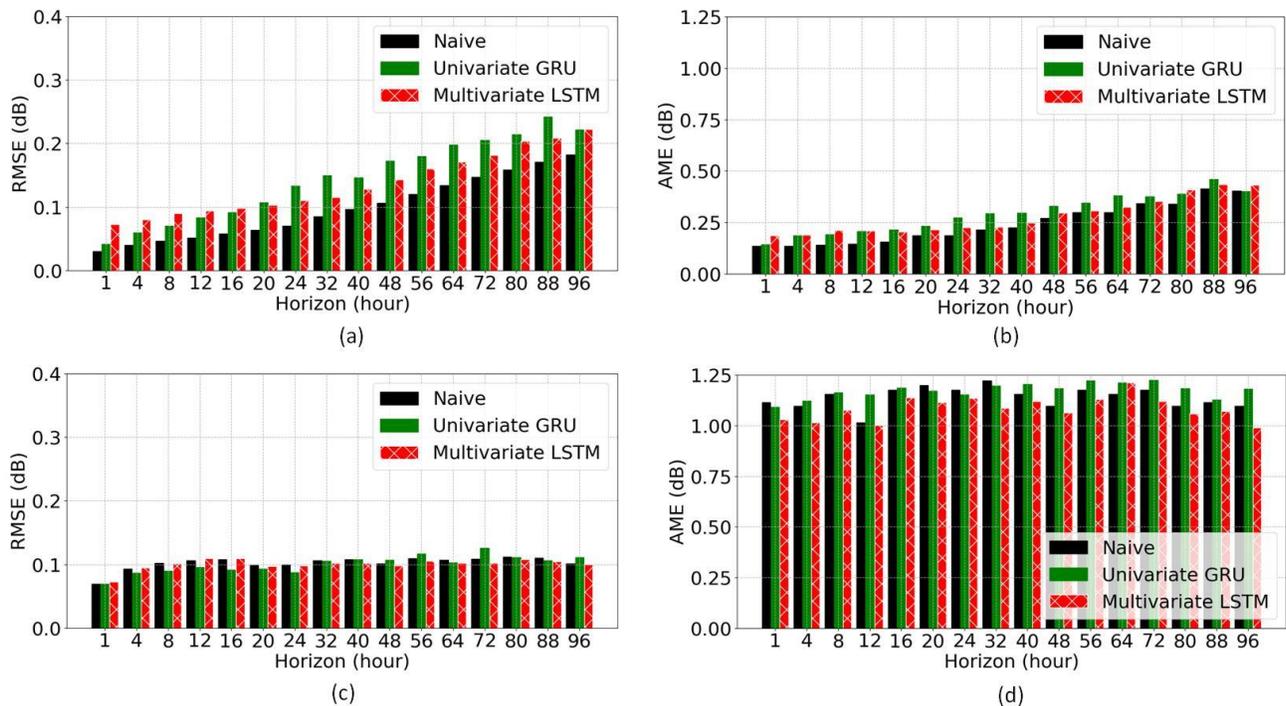


Fig. 10. Univariate GRU and multivariate LSTM-based transfer learning models. Performance evaluation using KB-3 (lightpath from opposite direction on the same route): (a) RMSE; (b) AME. Performance evaluation using KB-2 (lightpath in the same optical fiber on a portion of same route): (c) RMSE; (d) AME

from the source dataset.

B. Implementation details

The first step is the transfer of the source model to the target domain, as shown in Fig. 9. This involves using an SNR forecast model that has already been trained on a specific database (KB-1 in this case) to solve a new problem (SNR forecast of lightpath-2 and lightpath-3 in this case).

The forecast models used in the transfer learning demonstration are the best performing models, namely the univariate GRU model and the multivariate LSTM-based transfer learning model.

We adopt the parameter-based approach, which assumes that the structures learned by the two models trained on the source domain can be transferred to the two target domains.

This structure transfer is implemented by weight-sharing between the source (lightpath-1) and target (lightpath-2 and lightpath-3) domains.

The objective here is to perform a preliminary experiment with the transfer-learning approach to predict the SNR of two new lightpaths using a model that has been pre-trained on a different lightpath. Therefore, the weights to be used for models have not been optimized. The choice of weights could modify the performance of the model because the weight makes it possible to determine the interactions between all the neurons of each layer and the activation functions, and to define the knowledge acquired from the source domain [14].

The fine-tuning strategy in [26] was followed. Thus, the models were first trained using 24,445 instances in the source domain and then partially retrained in the target domain, as shown in Fig. 9. The target domain datasets were split into retraining (or fine-tuning) and test datasets according to a

80/20 ratio, as in [26]. This implied a fine-tuning process using 80% of the target domain datasets, i.e. 12,800 instances and 14,265 instances of the KB-2 and KB-3 databases, respectively. The performance of the models was evaluated using the remaining 20% of the target domain datasets, namely 3,200 and 3,567 instances of KB-2 and KB-3 datasets, respectively.

C. Performance results

The metrics used to evaluate the resulting models are the RMSE and AME. The results are shown in Fig. 10.

As shown by the RMSE curve in Fig. 10(a), the multivariate and univariate models using database KB-3 underperformed the naive model at all horizons, with an RMSE difference of up to 0.06 dB. Note that the target lightpath-3 and the source lightpath-1 are carried in different optical fibers and that the datasets have different distributions and variabilities. In contrast, as shown in Fig. 10(c), the multivariate LSTM-based transfer learning model using database KB-2 slightly outperformed the univariate GRU-based transfer learning and the naive models for all but the 24-hour forecast horizon.

The same observations can be made by looking at the AME curves in Fig. 10(b) and Fig. 10(d). First, for lightpath-3, we can see that the multivariate LSTM and univariate GRU transfer learning models using KB-3 cannot outperform the naive method from an AME perspective. However, for lightpath-2, the multivariate LSTM-based transfer learning model using KB-2 performs better than the naive model at all horizons, with an AME advantage of 0.11 dB, compared to 0.304 dB in the source KB-1 database, at the same 96-hour forecast horizon.

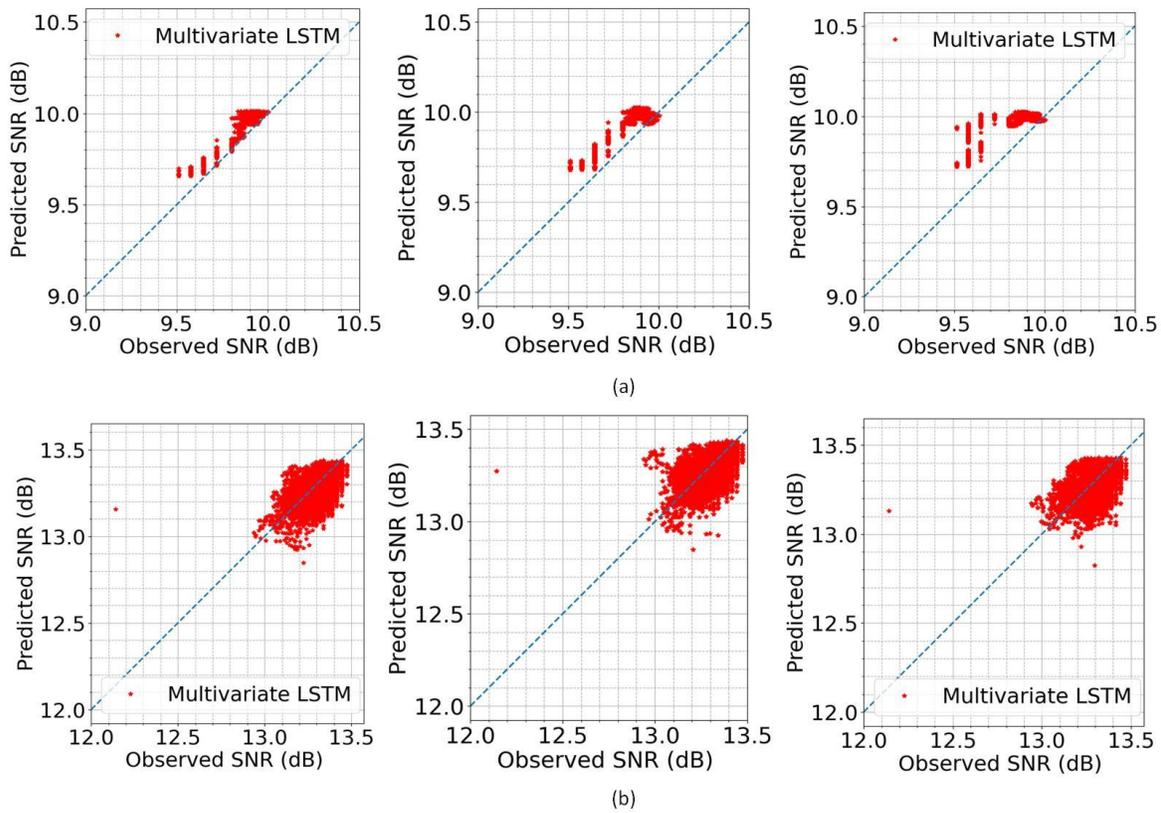


Fig. 11. Predicted SNR vs. observed SNR at 1-hour, 4-hour and 96-hour forecast horizons using the multivariate LSTM-based transfer learning model: (a) KB-3 (b) KB-2

Interestingly, the AME values for the lightpath-2 models are of the same order of magnitude as for the lightpath-1 models (in the 1.00-1.24 dB range), unlike those for lightpath-3, which are lower (0.20 – 0.49 dB), although the width of the SNR distribution for lightpath-2 is narrower than that for lightpath-1 and lightpath-3. This could potentially be explained by the similar range of SNR values in KB-2 and KB-1 (2.38 and 2.35 dB, respectively) and that they are almost twice the range of KB-3's SNR values (1.24 dB).

Finally, the models using transfer learning are faster to run than the classic training models for both KBs, with execution times of up to 53 ms for the models using transfer learning, compared to 320 ms for the basic models. Furthermore, the average training time for models using transfer learning is up to 25 minutes, compared to 75 minutes for the basic training models. The models were executed on a system with a 2.5 GHz Intel® Core™ i5-7200U processor, 8 GB RAM. Shorter training and execution times are precisely one of the benefits of transfer learning in cases where the databases are not sufficient to train and test the model. In such cases, the transfer learning approach can reduce the training time of the forecast models up to about 6 times and the calculation time of the forecast models up to 3 times.

Fig. 11 shows the predicted SNR vs. the observed SNR at the 1-hour, 4-hour and 96-hour forecast horizons for the multivariate LSTM-based transfer learning model. The plots for the models using KB-2 are more concentrated around the dotted line than those using KB-3, which shows a better forecast accuracy for the lightpath carried in the same optical fiber. Note that the choice of data for the target domain has an

impact on model performance. The more the target data has the same distribution and information as the source data, the better the performance of the model. When the approach of transferring the weights of the networks in the source domain is chosen, the target domains inherit source domain information. And this strategy produces satisfactory performance in scenarios where there is a similarity between source and target domains. In this study, the KB-2 target data comes from the same optical fiber as the KB-1 source data and the range of SNR values is approximately the same.

Overall, these preliminary results show the potential to reuse forecast models previously trained with data from one lightpath to predict the performance of other lightpaths carried in the same optical fibers through transfer learning. The performance of the multivariate LSTM-based transfer learning model is better than the naive method, with an advantage in terms of RMSE and AME.

VII. CONCLUSIONS

In this work, we proposed two multivariate LSTM and GRU models, referred to as LSTM-M-1 and GRU-M-1 respectively, to forecast the SNR of a 1300-km lightpath (lightpath-1). The historical performance metrics used were the pre-FEC BER, the channel received power and DGD, as well as the outside temperature and period of the day. The models were implemented, trained and tested using a set of 5 features over a 13-month observation period. The results show that the multivariate models performed slightly better than their univariate counterparts and a persistence model, with lower AME and higher R^2 scores for horizons lower than 96-

hour. In fact, even if the multivariate models are very close (in terms of RMSE performance), the multivariate LSTM-M-1 model stands out across all the short-term forecast horizons and exhibits the smallest maximum error values, with maximum error differences of up to 0.08 dB, 0.17 dB and 0.28 dB compared to the multivariate GRU-M-1, the univariate GRU and the univariate LSTM models, respectively.

In addition, a statistical analysis performed on the dataset for feature reduction revealed that the channel power was the best additional feature (amongst the 5 available features) to use along with historical SNR values for predicting future SNR over time horizons of up to 96 hours. The resulting 2-feature multivariate models underperformed the 5-feature models and the univariate models, but they outperformed the naive method and could be run in much shorter execution time than the 5-feature multivariate models and univariate models.

Finally, to explore the concept of transfer learning, the best performing multivariate LSTM forecast model built using KB-1 was fine-tuned using smaller datasets from a lightpath deployed in the same optical fiber on a shorter portion of the same route (lightpath-2) and from a lightpath carried in the optical fiber in the opposite direction on the same route (lightpath-3). It was possible to forecast the SNR for lightpath-2 carried in the same optical fiber as the source domain with an RMSE difference of up to 0.01 dB compared to the naive model, while maintaining an AME advantage of up to 0.13 dB. However, the LSTM model underformed the naive models across all forecast horizons for lightpath-3 carried in a different optical fiber. Based on these preliminary results, transfer learning could be a promising solution for short-term forecast horizons for lightpaths carried in the same optical fiber of the same route.

Overall, the best performance was achieved by using the 5-feature LSTM model. The model performed better than both the multivariate GRU and the univariate LSTM and GRU models in forecasting lightpath SNR to up to a 96-hour forecast horizon. Moreover, computation times for the multivariate models do not increase with the forecast horizon, contrary to the computation times for the univariate models, which increase as the forecast horizon increases.

The potential benefits of multivariate models for predicting lightpath performance has been demonstrated in this paper by using field datasets made available by a network operator. Further work will be required to validate the results using bigger PM datasets including additional features and network topology information. The growing AME benefit of the ML models with respect to the naive model as the forecast horizon lengthens also suggests that ML forecast models would deserve to be studied for horizons exceeding 4 days. Future work will also include automated optimization of hyperparameters at each forecast horizon to improve the performance of the models regardless of the dataset and horizon. Finally, in order to analyze the impact of the transfer learning on the SNR prediction, future work will focus on the comparison of conventional prediction models and models using several datasets from multiple sources.

REFERENCES

- [1] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of Transmission Estimation and Short Term Performance Forecast of Lightpaths," *Journal of Lightwave Technology*, vol. 38, pp. 2807-2814, 2020.
- [2] C. Tremblay, A. Michel, M. J. Tanoh, M. P. Bélanger, S. Clarke, D. W. Charlton, *et al.*, "Dynamics of polarization fluctuations in aerial and buried links," in *2017 19th International Conference on Transparent Optical Networks (ICTON)*, 2017, pp. 1-1.
- [3] A. Michel, "Monitoring of performance parameters in a metro ring using a 100G coherent transponder," M.Eng. Final Report, École de Technologie Supérieure (ÉTS), 2016.
- [4] Francesco Musumeci, Cristina Rottondi, Avishek Nag, Irene Macaluso, Darko Zibar, Marco Ruffini, *et al.*, "A Survey on Application of Machine Learning Techniques in Optical Networks," 2018.
- [5] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, *et al.*, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Optical Switching and Networking*, vol. 28, pp. 43-57, 2018/04/01/ 2018.
- [6] G. Choudhury, D. Lynch, G. Thakur, and S. Tse, "Two use cases of machine learning for SDN-enabled ip/optical networks: traffic matrix prediction and optical path performance prediction [Invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, pp. D52-D62, 2018.
- [7] R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, pp. D84-D99, 2018.
- [8] P. Samadi, D. Amar, C. Lepers, M. Lourdiane, and K. Bergman, "Quality of Transmission Prediction with Machine Learning for Dynamic Operation of Optical WDM Networks," in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1-3.
- [9] M. Abdel-Nasser, K. Mahmoud, O. A. Omer, M. Lehtonen, and D. Puig, "Link quality prediction in wireless community networks using deep recurrent neural networks," *Alexandria Engineering Journal*, vol. 59, pp. 3531-3543, 2020/10/01/ 2020.
- [10] C. Tremblay, S. Allogba, and S. Aladin, "Quality of Transmission Estimation and Performance Prediction of Lightpaths Using Machine Learning," presented at the 45th European Conf. Optical Commun. (ECOC), Dublin, Ireland, 2019.
- [11] S. Aladin, S. Allogba, A. V. S. Tran, and C. Tremblay, "Recurrent Neural Networks for Short-Term Forecast of Lightpath Performance," in *Optical Fiber Communication Conference (OFC) 2020*, San Diego, California, 2020, p. W2A.24.
- [12] C. Tremblay, "Towards Cognitive Management and Performance Monitoring in Coherent Optical Networks," presented at the CLEO: Science and Innovations, 2020.
- [13] A. Mezni, D. W. Charlton, C. Tremblay, and C. Desrosiers, "Deep Learning for Multi-Step Performance Prediction in Operational Optical Networks," in *Conference on Lasers and Electro-Optics*, Washington, DC, 2020, p. STh4M.1.
- [14] W. Mo, Y. Huang, S. Zhang, E. Ip, D. C. Kilper, Y. Aono, *et al.*, "ANN-Based Transfer Learning for QoT Prediction in Real-Time Mixed Line-Rate Systems," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1-3.
- [15] R. D. Marino, C. Rottondi, A. Giusti, and A. Bianco, "Assessment of Domain Adaptation Approaches for QoT Estimation in Optical Networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1-3.
- [16] J. Yu, W. Mo, Y. Huang, E. Ip, and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, pp. C48-C57, 2019.
- [17] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 1310-1315.
- [18] S. Allogba and C. Tremblay, "K-Nearest Neighbors Classifier for Field Bit Error Rate Data," in *2018 Asia Communications and Photonics Conference (ACP)*, 2018, pp. 1-3.
- [19] S. L. Woodward, L. E. Nelson, C. R. Schneider, L. A. Knox, M. O. Sullivan, C. Laperle, *et al.*, "Long-Term Observation of PMD and SOP on Installed Fiber Routes," *IEEE Photonics Technology Letters*, vol. 26, pp. 213-216, 2014.

- [20] E. e. C. c. Canada. (2011). *Données historiques - Climat - Environnement et Changement climatique Canada*. Available: https://climat.meteo.gc.ca/historical_data/search_historic_data_f.html
- [21] B. L. M. Yaméogo, D. W. Charlton, D. Doucet, C. Desrosiers, M. O. Sullivan, and C. Tremblay, "Trends in Optical Span Loss Detected Using the Time Series Decomposition Method," *Journal of Lightwave Technology*, vol. 38, pp. 5026-5035, 2020.
- [22] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 2nd edition ed. Melbourne, Australia: OTexts, 2018.
- [23] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu, "Memory in Memory: A Predictive Neural Network for Learning Higher-Order Non-Stationarity From Spatiotemporal Dynamics," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9146-9154.
- [24] H. Chouman, P. Djukic, C. Tremblay, and C. Desrosiers, "Forecasting Lightpath QoT with Deep Neural Networks," presented at the Optical Fiber Communication Conference (OFC) 2021, San Francisco, California, 2021.
- [25] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, p. 9, 2016/05/28 2016.
- [26] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, pp. 43-76, 2021.

held senior R&D and technology management positions for several organizations. As Engineering Manager at EXFO and Director of Engineering at Roptest, she was responsible for the development of fiber-optic test equipment. She also served as Product Manager at Nortel for DWDM systems. Her research interests include machine learning for optical networking applications, optical performance monitoring and filterless optical networking. She has been co-instructor for SC314 and SC210 hands-on courses of the Optical Society of America on optical fiber and polarization measurements.

Dr. Tremblay is a member of the Optical Society of America (OSA), the IEEE Photonics Society, as well as STARaCom and COPL Strategic Clusters of FRQNT. She is currently serving as Technical Program Committee (TPC) Member for OFC N3 Subcommittee "Architecture and software-defined control for metro and core networks" as well as Program Chair for the Photonics Networks and Devices (NETWORKS) OSA Topical Meeting.

Stéphanie Allogba received her B.Eng. degree in telecommunications from ESME Sudria, France, in 2011 and her M.A.Sc. degree in telecommunications networks engineering from École de technologie supérieure, Montréal, Canada, in 2015. Her Master's thesis focused on control and management algorithms for femtocell wireless networks. Since 2016, she is working towards a Ph.D. degree at École de technologie supérieure, with the Network Technology Lab. Her research interests include the application of machine learning algorithms for lightpath classification and performance prediction in optical networks.

Sandra Aladin received her B.Eng. degree in electronics from Université d'État d'Haiti, Port au Prince, Haiti, in 2009 and her M.A.Sc. degree in telecommunications network engineering from École de technologie supérieure, Montréal, QC, Canada, in 2018. Her Master's thesis focused on machine learning-based quality of transmission estimation in optical networks. Since 2018, she has been working as a Research Assistant on machine learning for optical networking with the Network Technology Lab, Department of Electrical Engineering, École de technologie supérieure, Montréal.

Christine Tremblay received her B.Sc. degree in Engineering Physics from Université Laval, Quebec City, Canada, in 1984, M.Sc. degree (Energy) from INRS-Énergie, Varennes, Canada, in 1985 and Ph.D. degree (Optoelectronics) from École Polytechnique de Montréal, Canada, in 1992. She is a Full Professor with the Department of Electrical Engineering and Associate Director for the Ph.D. Program at École de technologie supérieure. She is the Founding Researcher and Head of the Network Technology Lab. Before joining ÉTS, she was a Research Scientist with the National Optics Institute (INO) where she conducted research on integrated optical devices for communication and sensing applications. She