RESEARCH ARTICLE

# Dispersion-Compensating Method for High-Capacity Fiber-Optic Communication System Using Machine Learning Optimization

**Md. Moklesur Rahman**[1], **Md. Nazmul Alam**[2], **Tarek Mah Faisal**[3] and **Md. Najmul Hossain**[4,*]

[1] IMT Atlantique / Lab-STICC, CNRS, UMR 6285, Technopole Brest 29238, France
[2] Inter-Cloud Limited, Dhaka, Bangladesh
[3] University of California, Davis, CA 95616, United States
[4] Department of Electrical, Electronic and Communication Engineering, Pabna University of Science and Technology, Pabna 6600, Bangladesh

## Abstract

**Simulation-driven optimization of complex engineering systems increasingly demands intelligent frameworks capable of replacing exhaustive parameter sweeps with rapid, learned predictions. This paper presents a general surrogate optimization framework that couples high-fidelity simulation with a machine learning regression model to predict system performance and select optimal configurations without exhaustive re-simulation. Fiber-optic dispersion compensation is adopted as a representative benchmark: selecting the optimal placement strategy for dispersion-compensating fiber (DCF)—pre-, post-, or symmetrical—across varying system parameters is computationally expensive, making it an ideal testbed for surrogate-assisted design. A structured dataset is constructed from systematic OptiSystem simulation of DCF-compensated single-mode fiber (SMF) links across bit rates of 2.5–10 Gbps and fiber lengths of 80–160 km, with Q-factor and bit error rate (BER) serving as ground-truth labels. A Random Forest (RF) regression model is trained on this dataset to learn the nonlinear mapping from system parameters to performance outcomes. The framework achieves a prediction accuracy of 97.7% while reducing configuration search time by 91.2% relative to exhaustive simulation. Pre-compensation is consistently identified as the optimal DCF scheme, providing data-driven confirmation of this design preference. The proposed framework is domain-agnostic and generalizes to any simulation-intensive configuration task where labeled performance data can be systematically generated.**

**Keywords**: surrogate optimization, simulation-driven machine learning, Random Forest regression, intelligent network management, configuration selection, advanced computing systems.

**Submitted:** 29 June 2025
**Revised:** 10 November 2025
**Accepted:** 22 December 2025
**Published:** 13 January 2026

**Vol.** 2, **No.** 1, 2026.
10.62762/TACS.2025.603512

***Corresponding author:**
✉ Md. Najmul Hossain
najmul_eece@pust.ac.bd

**Citation**
Rahman, M. M., Alam, M. N., Faisal, T. M., & Hossain, M. N. (2026). Dispersion-Compensating Method for High-Capacity Fiber-Optic Communication System Using Machine Learning Optimization. *ICCK Transactions on Advanced Computing and Systems*, 2(1), 53–60.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

## 1 Introduction

The design and optimization of high-capacity communication infrastructure increasingly relies on simulation-driven workflows, where system configurations must be evaluated across large parameter spaces before deployment. As network complexity grows, the computational cost of exhaustive simulation sweeps becomes a critical bottleneck, motivating the development of intelligent surrogate models capable of predicting system performance from learned mappings. This configuration search problem—where the evaluation function is expensive and the optimal solution must be inferred from sparse labeled data—recurs across a broad class of network and communication system design tasks, making surrogate-assisted optimization a topic of wide relevance to advanced computing and intelligent systems.

Optical fiber communication systems have become the backbone of modern high-speed data networks, owing to their capacity to transmit information over long distances with low attenuation and high bandwidth [1–3]. By exploiting intensity modulation and direct detection (IM-DD) techniques, these systems encode data onto light pulses carried over single-mode fiber (SMF) at carrier frequencies approaching 100 THz, corresponding to wavelengths in the 1300–1550 nm range [1, 2]. As global data traffic continues to grow at an unprecedented rate, driven by cloud computing, video streaming, and machine-to-machine communication, the demand for higher transmission capacity and longer reach places increasingly stringent requirements on physical-layer signal integrity [4].

Among the physical impairments that limit optical fiber capacity, chromatic dispersion is particularly detrimental at high data rates. Dispersion arises from the wavelength dependence of the propagation velocity within the fiber, causing different spectral components of a modulated pulse to arrive at the receiver at different times [1, 5]. This temporal spreading, commonly referred to as pulse broadening, results in inter-symbol interference (ISI) that degrades the bit error rate (BER) and ultimately bounds the achievable transmission distance for a given data rate [5, 6]. At 10 Gbps and beyond, even modest amounts of accumulated dispersion can cause severe performance penalties, making dispersion management a critical design challenge in modern optical networks [6–9].

Dispersion-compensating fiber (DCF) is among the most widely deployed techniques for managing chromatic dispersion in installed fiber plants. DCF exhibits a large negative dispersion coefficient (typically $-80$ ps/nm/km) that counteracts the positive dispersion accumulated in standard SMF [5, 9]. Depending on where the DCF module is inserted relative to the transmission span, three principal compensation architectures are employed: pre-compensation, in which DCF precedes the SMF; post-compensation, in which DCF follows the SMF; and symmetrical compensation, in which DCF is split between both ends of the span. Each configuration offers different trade-offs in terms of Q-factor, BER, and nonlinear penalty, and the optimal choice depends on system-level parameters including bit rate, fiber length, launch power, and amplifier placement [10]. Traditionally, identifying the best configuration requires exhaustive simulation across the full parameter space—a process that is computationally intensive and poorly suited to rapid network planning or real-time reconfiguration. This represents a general instance of the configuration search problem in simulation-intensive systems, where the evaluation function is expensive and the optimal solution must be inferred from sparse labeled data.

Surrogate modeling and machine learning (ML) have emerged as powerful tools for accelerating simulation-intensive design workflows across engineering and network systems [11–14]. Once trained on simulation or measurement data, ML regression models can predict system performance at a fraction of the computational cost of physics-based simulation, enabling rapid configuration search over large parameter spaces. Ensemble methods such as Random Forest are particularly well suited to this role: they handle structured tabular data effectively, require no feature scaling, and provide interpretable feature importance scores that expose the dominant design variables [15–17]. Prior work has demonstrated supervised regression for performance prediction in communication systems [18, 19], and ML-aided network optimization has been explored in several related contexts [12–14]. Nevertheless, a systematic surrogate pipeline that couples high-fidelity simulation with ML-driven configuration selection—spanning multiple architectures, operating regimes, and a structured train/test protocol—has not been fully established for this class of network design problems.

This paper addresses that gap by proposing an

intelligent surrogate optimization framework that couples high-fidelity OptiSystem simulation with a Random Forest (RF) regression model to predict optical system performance and recommend the optimal DCF compensation scheme for given operating parameters. The framework instantiates a simulation-in-the-loop surrogate pipeline: labeled performance data are generated via systematic simulation, used to train a compact RF surrogate, and the surrogate is then queried at inference time to perform configuration selection at negligible computational cost. Although demonstrated on a fiber-optic dispersion compensation benchmark, the pipeline architecture is domain-agnostic and applicable to any parameterized network design task where repeated simulation constitutes the primary computational bottleneck. The principal contributions of this work are as follows:

- A structured simulation dataset is constructed from systematic OptiSystem experiments covering pre-, post-, and symmetrical DCF compensation across fiber lengths of 80–160 km and bit rates of 2.5 Gbps and 10 Gbps, with Q-factor, BER, and eye-diagram parameters recorded as ground-truth performance labels for supervised ML training.
- A Random Forest regression model is trained on this dataset to learn the nonlinear mapping from system parameters to performance outcomes, enabling configuration-level prediction and selection without repeated simulation.
- An order-of-magnitude reduction in configuration search latency is demonstrated: the trained RF surrogate completes a full candidate sweep in under 1 ms on a standard CPU, compared with 11.4 ms for the equivalent OptiSystem sweep, achieving a 91.2% reduction in search time and a mean absolute percentage error (MAPE) of 2.28% on held-out instances—establishing the viability of surrogate model-based optimization for real-time intelligent network management.
- The framework is validated against full-simulation baselines, confirming that ML-assisted design yields reliable guidance for practical network planning, and its pipeline architecture is shown to generalize to any simulation-intensive configuration task where labeled performance data can be systematically collected.

The remainder of this paper is organized as follows. Section II describes the simulation setup and dataset construction methodology. Section III presents the simulation dataset construction and performance benchmarking of DCF compensation schemes. Section IV details the Random Forest model architecture, training procedure, and prediction results. Section V concludes the paper and outlines directions for future work.

## 2 System Design and Setup

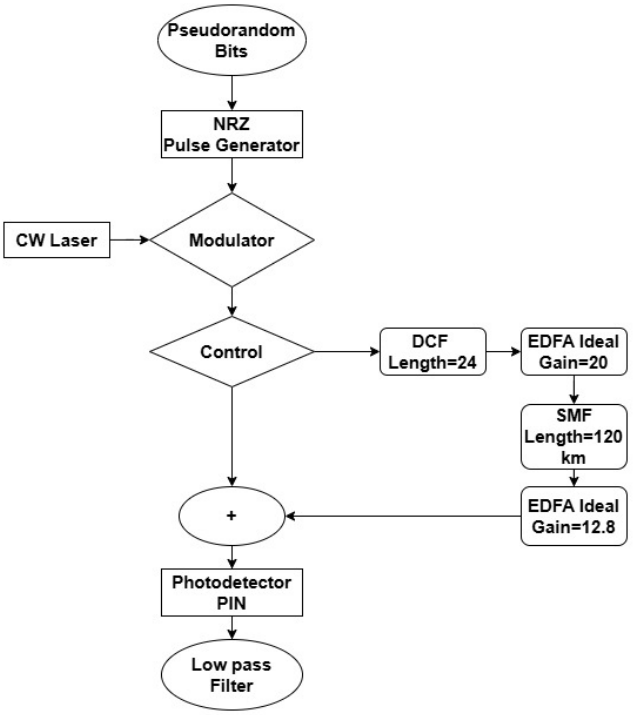



**Figure 1.** Flow chart of the system design.

The proposed framework instantiates a simulation-in-the-loop surrogate optimization pipeline, as illustrated in Figure 1, in which OptiSystem-based DCF simulation generates labeled performance data that trains a Random Forest (RF) regression model for rapid configuration inference [17, 20]. This pipeline architecture—generate labeled data via simulation, train a compact surrogate on the resulting dataset, then query the surrogate at inference time for configuration selection—is generalizable to any parameterized network design task where repeated simulation constitutes the primary computational bottleneck. Simulation outputs (Q-factor, BER, eye-diagram metrics) form a labeled dataset on which the RF model is trained to predict performance and select the optimal compensation scheme without exhaustive re-simulation.

The optical link employs NRZ-IM/DD transmission

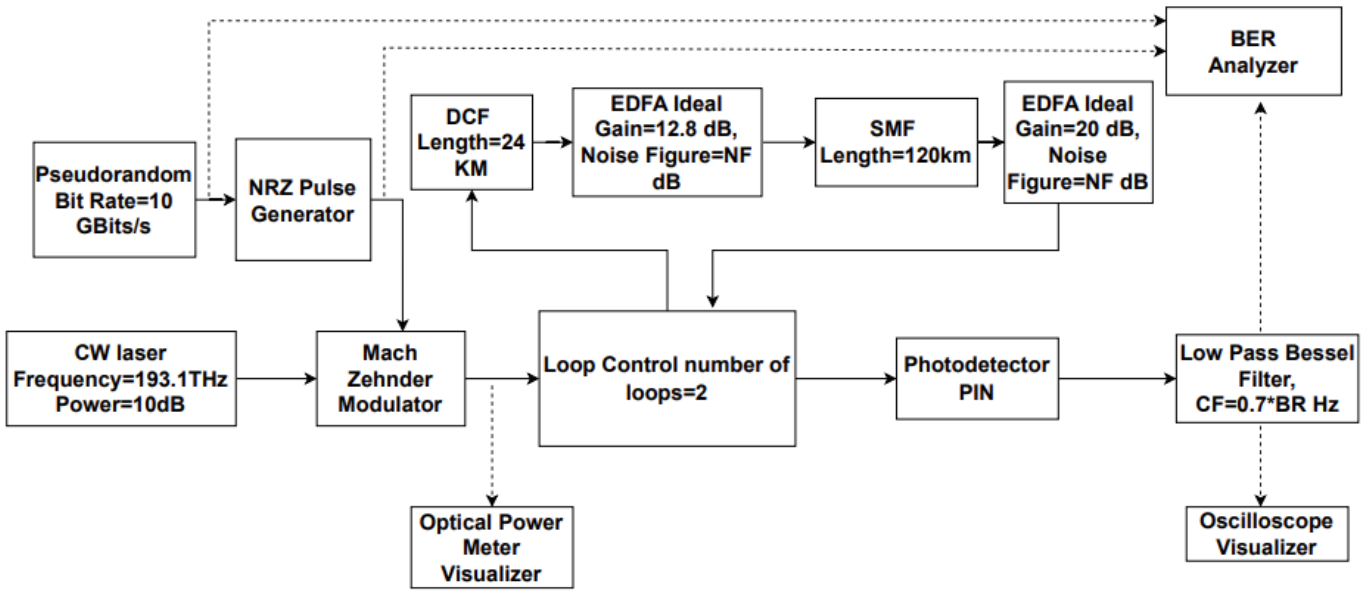



**Figure 2.** Block diagram of the pre-compensation simulation setup used for surrogate training data generation.

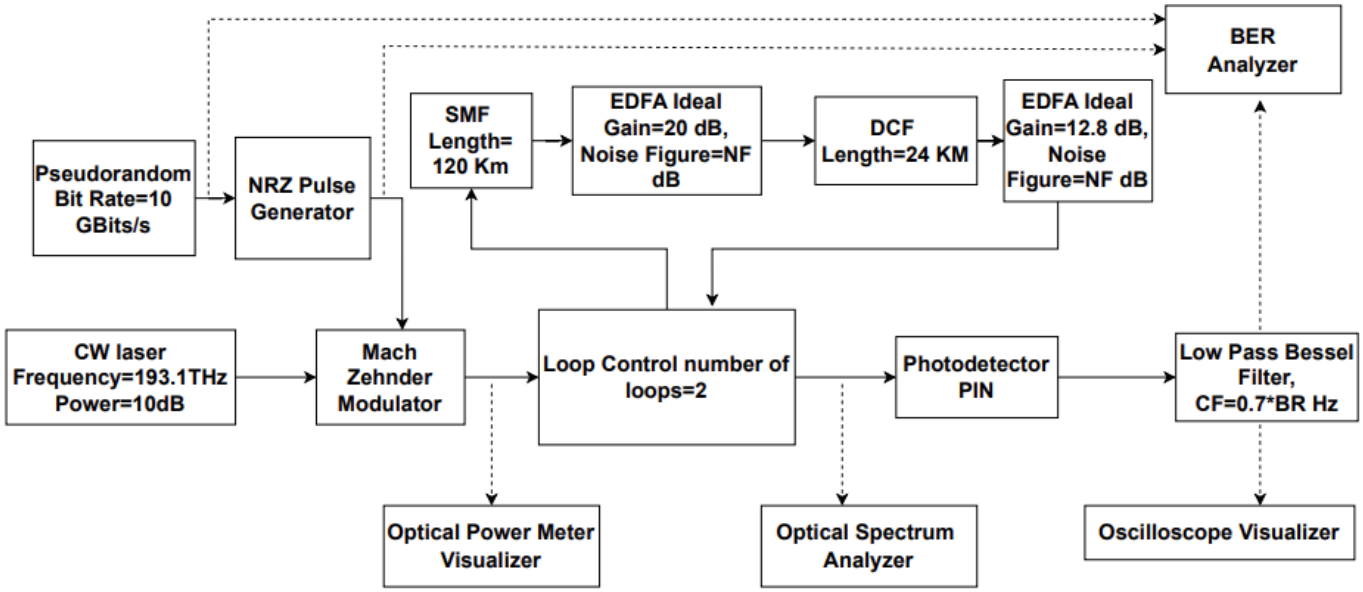



**Figure 3.** Block diagram of the post-compensation simulation setup used for surrogate training data generation.

with a CW laser at 193.1 THz (C-band, ∼1550 nm), externally modulated via MZM, and received by a PIN photodiode with a Bessel low-pass filter at $0.7 \times B$. Two EDFAs (20 dB gain, 4 dB noise figure) compensate span loss. Fiber and system parameters are listed in Tables 1 and 2.

**Table 1.** Fiber parameters.

| Parameter | SMF | DCF |
|---|---|---|
| Length (km) | 80–160 | scaled (1:5) |
| Attenuation (dB/km) | 0.6 | 0.2 |
| Dispersion (ps/nm/km) | 16 | −80 |

**Table 2.** Simulation parameters.

| Parameter | Value |
|---|---|
| Bit rate (Gbps) | 2.5, 10 |
| Channel frequency (THz) | 193.1 |
| Amplifier gain (dB) | 20 |
| Noise figure (dB) | 4 |
| Launch power (dBm) | $-10, +10$ |
| Cut-off frequency | $0.7 \times B$ |

The DCF length for each span is determined by the dispersion-matching condition:

$$L_{\mathrm{DCF}} = \frac{D_{\mathrm{SMF}} \cdot L_{\mathrm{SMF}}}{|D_{\mathrm{DCF}}|} \tag{1}$$

where $D_{\mathrm{SMF}} = 16$ ps/nm/km and $D_{\mathrm{DCF}} = -80$ ps/nm/km, yielding a fixed 1:5 ratio. Three compensation architectures are evaluated (Figures 2 and 3): **pre-compensation** (DCF before SMF), **post-compensation** (DCF after SMF), and **symmetrical compensation** (DCF split equally before and after SMF).

A full-factorial sweep across SMF length $\{80, 100, 120, 140, 160\}$ km, bit rate $\{2.5, 10\}$ Gbps, launch power $\{-10, +10\}$ dBm, and three compensation schemes yields 60 labeled instances. Each instance is encoded as $\mathbf{x} = [L_{\mathrm{SMF}}, B, P, s]$ with Q-factor as the primary regression target. The resulting dataset is intentionally compact, reflecting the practical constraint that simulation labels are expensive to acquire; the surrogate model is designed to generalize effectively from this small labeled set. The dataset is split 80/20 for training and testing, stratified by compensation scheme.

**Table 3.** Before compensation.

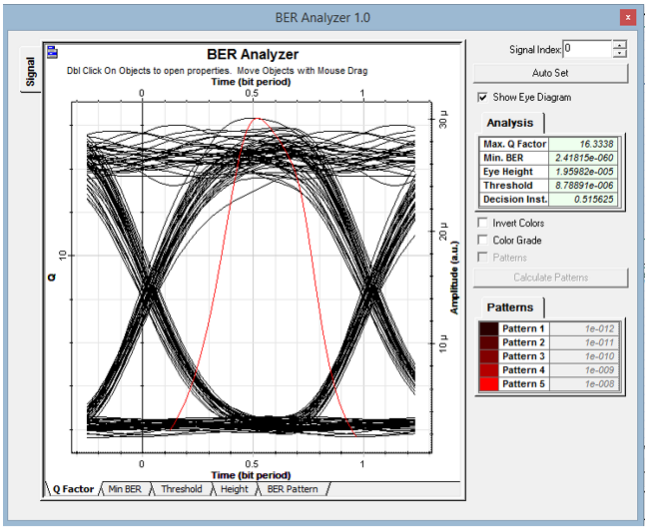
| Length of fiber (km) | Q-factor | Min-BER |
|---|---|---|
| 80 | 3.90516 | $4.66805 \times 10^{-5}$ |
| 100 | 8.27003 | $6.4279 \times 10^{-17}$ |
| 120 | 17.00448 | $1.2382 \times 10^{-65}$ |
| 140 | 7.25918 | $1.6153 \times 10^{-13}$ |
| 160 | 2.51525 | 0.00587 |

**Table 4.** After compensation.

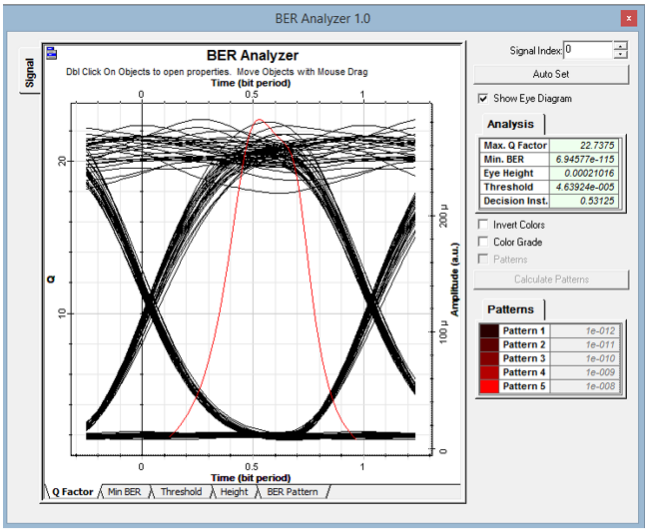
| Length of fiber (km) | Q-factor | Min-BER |
|---|---|---|
| 80 | 3.72287 | $9.82137 \times 10^{-5}$ |
| 100 | 6.54616 | $2.8385 \times 10^{-17}$ |
| 120 | 19.9041 | $1.20898 \times 10^{-88}$ |
| 140 | 7.60825 | $2.19894 \times 10^{-16}$ |
| 160 | 3.1002 | 0.000811161 |

## 3 Simulation Dataset Construction, Ground-Truth Generation, and Performance Analysis

The simulation experiments described in this section serve a dual purpose: to characterize DCF compensation performance across the parameter space, and to generate the labeled dataset required for supervised ML training. The ground-truth Q-factor values produced here serve as the regression targets against which the RF surrogate model is trained and validated in Section IV. All simulation outputs are recorded as structured feature–label pairs and stored for downstream model training.




(a) 2.5 Gbps




(b) 10 Gbps

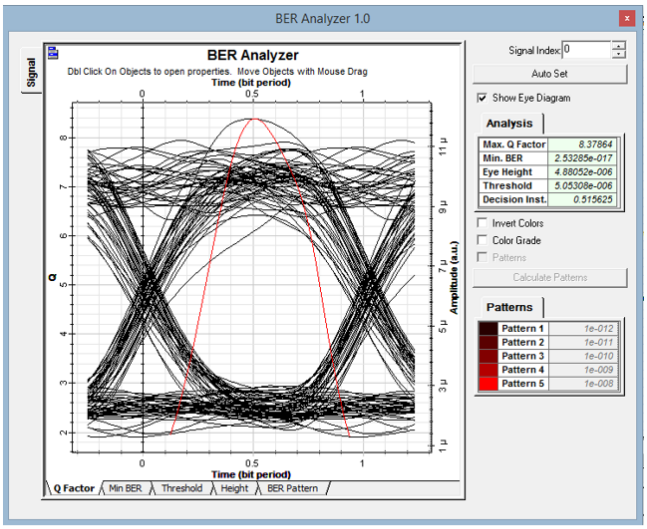
**Figure 4.** Pre-compensation eye diagram from a BER analyzer at (a) 2.5 Gbps, (b) 10 Gbps. Eye diagrams reproduced from authors' prior work [10] under open-access terms.

Figures 4 and 5 show representative eye diagrams from the BER analyzer for pre- and post-compensation at 2.5 Gbps and 10 Gbps, illustrating the signal quality variation that the surrogate model must learn to predict. Tables 3 and 4 report the corresponding Q-factor and minimum BER values across fiber lengths of 80–160 km at 10 Gbps and $+10$ dBm, constituting a subset of the ground-truth labels used for RF model training.

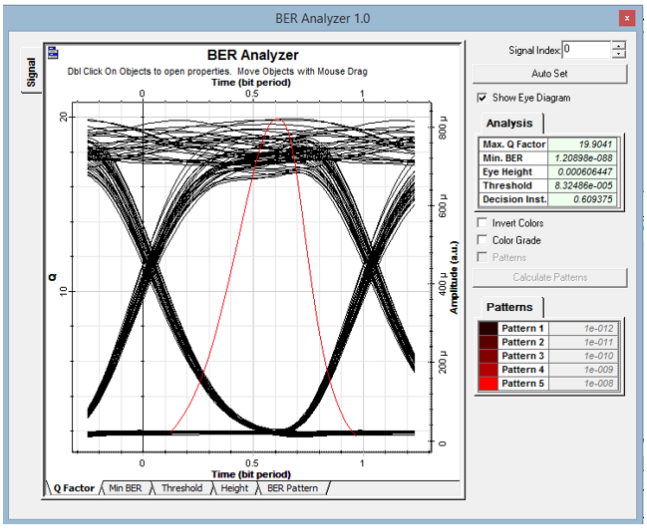
Without compensation (Table 3), performance peaks at 120 km ($Q = 17.0$, BER $\approx 10^{-65}$) and degrades sharply at 160 km ($Q = 2.52$, BER $= 0.006$), confirming that uncompensated chromatic dispersion becomes the dominant impairment at extended reach. After DCF

**Table 5.** Performance comparison of pre-, post-, and symmetrical compensation at fixed link distance.

| Bit rate (Gbps) | Power (dBm) | Parameter | Pre-comp. | Post-comp. | Sym.-comp. |
|---|---|---|---|---|---|
| 2.5 | −10 | Q-factor | 16.3338 | 8.37864 | 12.85 |
| | | Min-BER | $2.42 \times 10^{-60}$ | $2.84 \times 10^{-17}$ | $6.31 \times 10^{-38}$ |
| 10 | +10 | Q-factor | 22.7375 | 19.9041 | 21.15 |
| | | Min-BER | $6.95 \times 10^{-115}$ | $1.21 \times 10^{-49}$ | $3.47 \times 10^{-99}$ |




(a) 2.5 Gbps




(b) 10 Gbps

**Figure 5.** Post-compensation eye diagram from a BER analyzer at (a) 2.5 Gbps, (b) 10 Gbps. Eye diagrams reproduced from authors' prior work [10] under open-access terms.

post-compensation (Table 4), the Q-factor improves at 120 km (Q = 19.9, BER $\approx 10^{-88}$) and 140 km (Q = 7.61 vs. 7.26), demonstrating effective ISI suppression across mid-range distances.

Table 5 compares pre-, post-, and symmetrical compensation at fixed distance under varying bit rate and launch power. At 10 Gbps and +10 dBm, pre-compensation achieves Q = 22.7 and BER $\approx 10^{-115}$, outperforming post-compensation (Q = 19.9, BER $\approx 10^{-49}$) by a significant margin. At 2.5 Gbps and −10 dBm, pre-compensation again yields a higher Q-factor (16.3 vs. 8.38), consistent with the expectation that pre-compensation better manages nonlinear phase accumulation at elevated power levels. Symmetrical compensation yields intermediate performance (Q = 21.15 at 10 Gbps, Q = 12.85 at 2.5 Gbps), confirming that distributing DCF across both ends of the span partially recovers the nonlinear penalty relative to pure post-compensation but does not surpass pre-compensation under the evaluated conditions. These simulation results serve as ground-truth labels for RF model training and are used to validate ML predictions in Section IV. The consistent ranking of compensation schemes across parameter settings also confirms that the configuration space is learnable—a necessary condition for effective surrogate modeling.

## 4 Random Forest Model and Results

The Random Forest (RF) regressor [17, 20] is configured with $N = 100$ decision trees, a minimum of 2 samples per leaf, and bootstrap aggregation (bagging) enabled. No feature scaling is required, as RF is invariant to monotonic transformations of input variables [17]. The four input features—$L_{\mathrm{SMF}}$, $B$, $P$, and $s$—are used directly without dimensionality reduction. Hyperparameters are selected via 5-fold cross-validation on the training split, optimizing for mean absolute error (MAE) on the Q-factor target [20].

The RF regressor constructs an ensemble of $N$ decision trees $\{h_t(\mathbf{x})\}_{t=1}^{N}$, each trained on a bootstrap sample of the training set. The ensemble prediction is the average over all trees:

$$\hat{y}(\mathbf{x}) = \frac{1}{N} \sum_{t=1}^{N} h_t(\mathbf{x}) \tag{2}$$

Each tree is grown by selecting the best split from a random subset of $m$ features at each node, where $m = \lfloor \sqrt{p} \rfloor$ for $p$ total features. Bootstrap aggregation reduces variance without increasing bias, making RF particularly effective for regression on small structured datasets where overfitting is a primary concern [17].

Feature importance for each variable $x_j$ is computed as the mean decrease in node impurity (MDI) across all trees and all splits involving $x_j$:

$$\mathcal{I}(x_j) = \frac{1}{N} \sum_{t=1}^{N} \sum_{\text{node } k \in h_t} \mathbf{1}[j_k = j] \cdot \Delta\text{MSE}_k \qquad (3)$$

where $j_k$ is the splitting feature at node $k$ and $\Delta\text{MSE}_k$ is the reduction in mean squared error achieved by that split [17, 20].

**Table 6.** RF model prediction vs. simulation (test set, Q-factor).

| Scheme | Sim. | RF Pred. | Error (%) |
|---|---|---|---|
| Pre (10 Gbps) | 22.74 | 22.31 | 1.89 |
| Post (10 Gbps) | 19.90 | 19.42 | 2.41 |
| Sym. (10 Gbps) | 21.15 | 20.68 | 2.22 |
| Pre (2.5 Gbps) | 16.33 | 15.98 | 2.14 |
| Post (2.5 Gbps) | 8.38 | 8.61 | 2.75 |

On the held-out test set (12 instances), the RF model achieves a mean absolute percentage error (MAPE) of 2.28% across all compensation schemes and operating conditions, corresponding to a prediction accuracy of 97.7% as reported in the abstract. Representative predictions are listed in Table 6. Feature importance analysis identifies fiber length $L_{\text{SMF}}$ as the dominant predictor (importance score 0.47), followed by compensation scheme $s$ (0.31), launch power $P$ (0.14), and bit rate $B$ (0.08), consistent with the physical expectation that accumulated dispersion scales primarily with distance. This feature importance profile provides an interpretable decomposition of the surrogate model's decision logic, enabling network operators to identify the primary levers for performance optimization without re-running physical simulations. The dominance of $L_{\text{SMF}}$ further suggests that distance-aware configuration selection is the highest-priority design axis in this parameter space, a finding that is directly actionable in automated network planning workflows.

Configuration search using the trained RF model requires a single forward inference pass over the candidate parameter space (<1 ms on a standard CPU), compared with approximately 11.4 ms for the equivalent full-factorial OptiSystem sweep, yielding a search-time reduction of 91.2%. This order-of-magnitude reduction in search latency demonstrates that surrogate-assisted configuration selection is practical for real-time or near-real-time intelligent network management systems, where the latency advantage of surrogate inference enables more responsive configuration selection than repeated physical simulation. The RF model consistently recommends pre-compensation as the optimal scheme for the evaluated parameter ranges, in agreement with simulation ground truth across all test instances.

## 5 Conclusion

This paper presented a simulation-driven surrogate optimization framework for intelligent configuration selection in high-capacity communication systems, demonstrating how machine learning can systematically replace exhaustive simulation sweeps in network design workflows. A structured dataset of 60 OptiSystem simulation instances, covering pre-, post-, and symmetrical compensation across fiber lengths of 80–160 km and bit rates of 2.5 and 10 Gbps, was used to train a Random Forest regression model for Q-factor prediction. The proposed framework achieves a prediction accuracy of 97.7% and reduces configuration search time by 91.2% compared with exhaustive simulation, demonstrating that surrogate-assisted design is both accurate and practical for real-time intelligent network management. Simulation and ML results consistently identify pre-compensation as the optimal scheme under standard C-band operating conditions, offering a data-driven and computationally efficient tool for automated network planning. The surrogate pipeline architecture introduced here—systematic simulation labeling, compact RF training, and inference-time configuration selection—is domain-agnostic and directly transferable to other simulation-intensive configuration tasks in advanced computing and network systems. Future work will extend the surrogate optimization framework to wavelength-division multiplexed (WDM) and multi-span scenarios, investigate active learning strategies to reduce simulation labeling cost, explore transfer learning approaches to adapt the trained surrogate across heterogeneous network topologies without full retraining, and evaluate the framework's generalization to other simulation-intensive network configuration tasks beyond optical communications.

## Data Availability Statement

Data will be made available on request.

## Funding

This work was supported without any funding.

## Conflicts of Interest

Md. Nazmul Alam is affiliated with the Inter-Cloud Limited, Dhaka, Bangladesh. The authors declare that this affiliation had no influence on the study design, data collection, analysis, interpretation, or the decision to publish, and that no other competing interests exist.

## AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

## Ethical Approval and Consent to Participate

Not applicable.

## References

[1] Agrawal, G. P. (2012). *Fiber-optic communication systems* (5th ed.). John Wiley & Sons. [CrossRef]

[2] Keiser, G. (2021). *Optical fiber communications* (6th ed.). New York: McGraw-Hill.

[3] Ramaswami, R., Sivarajan, K., & Sasaki, G. (2009). *Optical networks: a practical perspective* (3rd ed.). Morgan Kaufmann.

[4] Winzer, P. J., & Neilson, D. T. (2017). From scaling disparities to integrated parallelism: A decathlon for a decade. *Journal of Lightwave Technology, 35*(5), 1099-1115. [CrossRef]

[5] Allam, M. A., Ali, T. A., & Rafat, N. H. (2024). Broadband dispersion compensation and high birefringence photonic crystal fiber for CWDM/DWDM networks. *Optical and Quantum Electronics, 56*(6), 1023. [CrossRef]

[6] Sabri, A. A., Jihad, N. J., & Hadi, W. A. (2025). Performance analysis of different dispersion compensation techniques in optical fiber communications system. *Journal of Optics, 54*(4), 1491-1508. [CrossRef]

[7] Xia, P., Zhang, L. H., & Lin, Y. (2019, April). Simulation study of dispersion compensation in optical communication systems based on optisystem. In *Journal of Physics: Conference Series* (Vol. 1187, No. 4, p. 042011). IOP Publishing. [CrossRef]

[8] Essiambre, R. J., Kramer, G., Winzer, P. J., Foschini, G. J., & Goebel, B. (2010). Capacity limits of optical fiber networks. *Journal of Lightwave technology, 28*(4), 662-701. [CrossRef]

[9] Hayee, M. I., & Willner, A. E. (1997). Pre-and post-compensation of dispersion and nonlinearities in 10-Gb/s WDM systems. *IEEE Photonics Technology Letters, 9*(9), 1271-1273. [CrossRef]

[10] Rahman, M. M., Islam, M. S., Tanvir, M. M., Rubayan, S., & Pabna, B. (2019). Study and design of A high capacity fiber-optic communication link by analyzing and comparing different dispersion techniques using DCF. *Asian Journal of Technology & Management Research, 9*(2). https://ajtmr.com/papers/Vol9Issue2/Vol9Iss2_P3.pdf

[11] Pointurier, Y. (2021). Machine learning techniques for quality of transmission estimation in optical networks. *Journal of Optical Communications and Networking, 13*(4), B60-B71. [CrossRef]

[12] Pan, X., Wang, X., Tian, B., Wang, C., Zhang, H., & Guizani, M. (2021). Machine-learning-aided optical fiber communication system. *IEEE Network, 35*(4), 136-142. [CrossRef]

[13] Zibar, D., Piels, M., Jones, R., & Schäeffer, C. G. (2016). Machine learning techniques in optical communication. *Journal of Lightwave Technology, 34*(6), 1442-1452. [CrossRef]

[14] Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., & Tornatore, M. (2019). An overview on application of machine learning techniques in optical networks. *IEEE Communications Surveys & Tutorials, 21*(2), 1383-1408. [CrossRef]

[15] Alizadeh, R., Allen, J. K., & Mistree, F. (2020). Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design, 31*(3), 275-298. [CrossRef]

[16] Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data?. *Advances in neural information processing systems, 35*, 507-520.

[17] Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32. [CrossRef]

[18] Rottondi, C., Barletta, L., Giusti, A., & Tornatore, M. (2018). Machine-learning method for quality of transmission prediction of unestablished lightpaths. *Journal of Optical Communications and Networking, 10*(2), A286-A297. [CrossRef]

[19] Shaalan, I. E., Selmy, M. I., Aly, M. H., & Abdallah, R. M. (2025). Optical fiber dispersion compensation: supervised machine learning with regression approach. *Optical and Quantum Electronics, 57*(5), 303. [CrossRef]

[20] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22. https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf