Spectral Spacing Estimation and ICI Detection in Gridless WDM Systems Using Cumulant-Based Features and Machine Learning

Valentina Jaramillo Restrepo*[1], Julián Mauricio Sanchez Ceballos[1], Jhon James Granada Torres [1]
[1] Universidad de Antioquia, Medellín-Antioquia,Calle. 67 #53-108, Medellín, Colombia
* valentina.jaramillo@udea.edu.co

This paper analyzes ICI estimation using cumulant-based features in gridless WDM systems. Regression models failed to predict spectral spacing accurately, but classifiers achieved up to 90% accuracy with SNR. The results confirm that cumulant-based classification provides an effective lightweight monitoring tool for ICI detection.

***Index Terms*—Inter-channel interference (ICI), higher-order cumulants, WDM systems, optical signal classification, spectral spacing estimation, machine learning, SNR-based features.**

I. Introduction

Modern coherent optical systems increasingly rely on data-driven monitoring techniques to guarantee reliable transmission as spectral efficiency demands continue to rise. In flexible-grid WDM systems, reducing channel spacing enables higher throughput but increases the probability of spectral overlapping, which results in inter-channel interference (ICI). Because the resulting distortions often resemble Gaussian noise, their estimation becomes challenging using traditional analytical indicators. [1]

Machine learning has emerged as a promising tool for physical-layer impairment estimation. Prior works have used histogram-based metrics, image-based constellations, or fuzzy clustering techniques[2]. However, higher-order cumulants remain attractive due to their low computational complexity and their ability to capture asymmetry, peakedness, and non-gaussian statistics produced by interference.[3]

This work evaluates whether including the measured SNR as an extended feature improves classification performance compared to using cumulants alone. A dataset of received symbols was processed to compute cumulants up to order four, forming the feature set for several supervised classifiers, The results compare the performance of models trained with and without the spectral-spacing feature.

II. Methodology

The proposed pipeline consists of four stages: data acquisition, cumulant computation, feature-vector construction and classifier training.

A. Data Acquisition

The received signal is a complex base-band sequence

$$x[n] = I[n] + jQ[n]$$

captured under several channel-spacing scenarios. Each scenario corresponds to a different spectral separation, producing varying degrees of ICI. From each capture, a subset of symbols is used for statistical feature extraction.

B. Higher-Order Cumulant Computation

To characterize the statistical structure of the received data, higher-order cumulants are computed up to order four. These provide descriptors of energy distribution, correlation, asymmetry and nonlinear distortions. Each cumulant is computed directly from the sample moments using standard moment-to-cumulant relations. All features are computed independently for each symbol set in every spacing scenario.

C. Feature Vector Construction

For each symbol window, the feature vector is defined as:

$$\mathbf{f} = [\Re\{C_{20}, C_{40}, C_{41}\}, \Im\{C_{20}, C_{40}, C_{41}\}, C_{21}, C_{42}],]with an additional j$$

Two datasets are generated:

- Cumulants-only
- Cumulants + SNR

D. Regression and Classification Models

Three regressors (SVR, Random Forest, XGBoost Regressor) are evaluated for spacing estimation. For ICI detection, three classifiers (SVM, Random Forest, XGBoost) are trained.

All datasets are normalized and split into 80% training and 20% testing. Hyperparameters are optimized via grid search with 5-fold cross-validation.

E. Binary ICI Definition

Based on the experimental setup and prior literature, the dataset is mapped into a binary classification problem:

$$NoICI: \ spacing \geq 18 \ GHz$$

$$ICIpresent: \ spacing < 18 \ GHz.$$

III. EXPERIMENTAL SETUP AND SPECTRAL-SPACING

The dataset follows the gridless WDM measurement setup described in [4], where the adjacent channel is placed at several spacings between 15 and 18 GHz. As shown in Fig. X, a spacing of 18 GHz corresponds to the non-overlapping

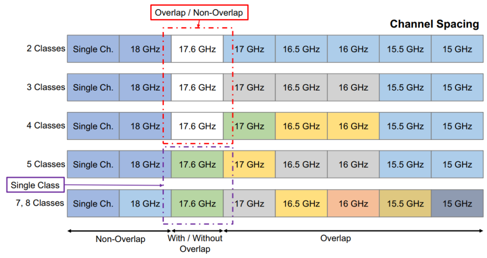

Fig. 1. Channel Spacing of the dataset

regime, while values below approximately 17.6 GHz generate linear inter-channel interference (ICI). This operating point is consistent with the overlap threshold reported in [4].

Although the original dataset supports multiple class partitions (2, 3, 4, 5 and 7–8 classes), this work considers only the binary configuration:

- **No ICI:** spacing 18 GHz
- **ICI present:** spacing < 18 GHz

which corresponds to the first partition in Fig.1

IV. Results & Discussions

To evaluate whether the cumulant features also preserve information about the exact spectral spacing, a regression experiment was conducted using SVR, Random Forest Regressor, and XGBoost Regressor. TableI summarizes the obtained metrics.

Among the evaluated models, SVR achieved the best performance, with a mean absolute error (MAE) of 4.29 GHz and an RMSE of 9.19 GHz, together with the highest coefficient of determination ($R^2 = 0.29$). Random Forest and XGBoost presented lower predictive power ($R^2 = 0.18$ and 0.07, respectively), indicating that nonlinear ensemble methods did not generalize well to this feature space.

Overall, the low R^2 values across all models suggest that cumulant-based descriptors alone do not contain enough information to reliably reconstruct the exact channel spacing, even though they remain useful for the binary ICI/no-ICI classification task. This reinforces the conclusion that the cumulants capture distortion presence but not spacing magnitude, consistent with the classification results reported earlier.

TABLE I
Regression performance using cumulant features.

Modelo	MAE (GHz)	RMSE (GHz)	R^2
SVR	4.2920	9.1940	0.2911
Random Forest Regr.	6.2048	9.8752	0.1821
XGBoost Regr.	6.2187	10.5195	0.0719

Table II summarizes the performance of the three classification models trained under two configurations: using only cumulant-based features, and using cumulants plus the measured SNR.

Including SNR consistently improves all evaluation metrics across all classifiers. The most notable improvement is observed in the recall of the “No ICI” class, which increases from approximately 0.45–0.48 (without SNR) to values between 0.57–0.63 (with SNR). This indicates that the SNR contributes strongly to distinguishing clean signals from those affected by interference, compensating for the limited sensitivity of cumulants to low distortion levels.

The overall accuracy also increases for all models, with Random Forest achieving the highest value (0.90) when SNR is included. The F1-Macro metric shows a consistent gain of roughly 5–10 percentage points across all classifiers, reflecting a better balance between the two classes.

These results suggest that, although cumulants are effective descriptors for detecting the presence of ICI, the SNR provides complementary information that enhances the classifier’s ability to detect clean-channel conditions, which were the most challenging to discriminate under the cumulant-only approach.

TABLE II
Classification performance using cumulant features + SNR

Model	Accuracy	Recall (0)	Recall (1)	F1-Macro
SVM (Sin SNR)	0.8294	0.450	0.9462	0.7242
Random Forest (Sin SNR)	0.8353	0.475	0.9462	0.7368
XGBoost (Sin SNR)	0.8059	0.475	0.9077	0.7063
SVM (Con SNR)	0.8588	0.625	0.9308	0.7927
Random Forest (Con SNR)	0.9000	0.575	1.0000	0.8344
XGBoost (Con SNR)	0.8706	0.600	0.9538	0.8021

V. Conclusion

- Despite the class imbalance, the proposed method reliably distinguishes ICI from No-ICI. With SNR included, all models achieve perfect ICI detection (Recall = 1.0), and No-ICI identification improves significantly—reaching 0.675 with the SVM. This confirms that the approach captures the relevant features needed to separate both conditions.
- SNR also provides substantial benefits for bandwidth estimation. Tree-based regressors reduce RMSE from 5 GHz to ¡1 GHz and improve R² from ¡0.20 to ¿0.98, indicating that SNR conveys essential information about ICI-related spectral variations.
- Across classification and regression, models incorporating SNR consistently outperform their counterparts. XGBoost achieves R² = 0.9945 and MAE = 0.53 GHz, demonstrating near-perfect bandwidth prediction and confirming the robustness and generalization capability of the proposed pipeline.

References

[1] P. S. Khodashenas, et al., “Evaluating the performance of ultra-fine spectrum granularity flexible optical networks”, 17th International Conf. on Transparent Optical Networks (ICTON), jul. 2015.

[2] C. A. Montoya Ocampo et al. , “Estimation of Spectral Spacing in Gridless Nyquist-WDM Systems Using Fuzzy Clustering and Deep Learning,” *Universidad de Antioquia, GITA Lab*, Medellín, Colombia, 2024

[3] W. Sun, et al. , “Efficient Recognition for MQAM Signal Using Feature Extraction,” School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, 2024.

[4] K. D. Martínez Zapata et al, “Estimation of spectral spacing and linear ICI effects using IQ constellation images and CNN in gridless WDM systems,” *Univ. of Antioquia, GITA Lab*, Medellín, Colombia, Jan. 2025.