Time Series Forecastability Measures

Rui Wang
Amazon Web Services
rwngamz@amazon.com

Steven Klee
Amazon Web Services
sklee@amazon.com

Alexis Roos
Amazon Web Services
alexiroo@amazon.com

Abstract

This paper proposes using two metrics to quantify the forecastability of time series prior to model development: the spectral predictability score and the largest Lyapunov exponent. Unlike traditional model evaluation metrics, these measures assess the inherent forecastability characteristics of the data before any forecast attempts. The spectral predictability score evaluates the strength and regularity of frequency components in the time series, whereas the Lyapunov exponents quantify the chaos and stability of the system generating the data. We evaluated the effectiveness of these metrics on both synthetic and real-world time series from the M5 forecast competition dataset. Our results demonstrate that these two metrics can correctly reflect the inherent forecastability of a time series and have a strong correlation with the actual forecast performance of various models. By understanding the inherent forecastability of time series before model training, practitioners can focus their planning efforts on products and supply chain levels that are more forecastable, while setting appropriate expectations or seeking alternative strategies for products with limited forecastability.

ACM Reference Format:
Rui Wang, Steven Klee, and Alexis Roos. 2025. Time Series Forecastability Measures. In *Proceedings of the 1st Workshop on "AI for Supply Chain: Today and Future" @ 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/XXXXXX.XXXXXX

1 Introduction

In the rapidly evolving landscape of supply chain management, accurate time series forecasting has become an indispensable tool for demand prediction, inventory optimization, and supply planning [3, 8–10, 15]. However, the effectiveness of these forecasts is intrinsically tied to the inherent forecastability of the underlying data. Not all time series exhibit the same degree of forecastability, and this variability can significantly impact the reliability of business decisions based on these predictions.

Traditionally, practitioners assess forecastability post hoc—by training models and evaluating performance. Although effective, this process is computationally expensive and can lead to wasted effort in inherently unpredictable series. We propose a more systematic alternative: using spectral predictability [7] and Lyapunov exponents [4] to quantify a time series' forecastability a priori [17]. We will demonstrate how these metrics can be systematically applied to time series data to identify the inherent difficulty of forecasting tasks and support better planning and resource allocation.

The spectral predictability evaluates the strength and complexity of frequency components within a time series, providing insights into its underlying patterns and cyclicality. Lyapunov exponent analysis, on the other hand, measures the stability and chaos of the data-generating system, offering insight into long-term behavior. Together, they offer complementary views into a series' structure and long-term dynamics.

This approach is particularly useful in supply chain management [2, 12, 14], where data is highly heterogeneous between products, categories, and regions. By understanding the forecastability of time series at various aggregation levels—such as individual products, product categories, or regional sales—decision-makers can better navigate complex networks, focus modeling efforts on more predictable areas, allocate resources efficiently, and set realistic expectations for forecasting performance.

We validate the use of these metrics through experiments on both synthetic and real-world datasets. In synthetic data, we show that spectral predictability and Lyapunov exponents strongly correlate with the underlying complexity of time series, effectively distinguishing between simple, noisy, chaotic, and random signals. In the hierarchical M5 dataset, we observed strong correlations between forecastability scores and actual forecast performance at different aggregation levels. Together, these findings demonstrate that the proposed use of these metrics offers a practical and computationally efficient way to assess time-series forecastability and guide forecasting strategies. They can set expectations on forecast performance and potentially inform hedging or intervention strategies, such as using different models for items with different levels of forecastability. Furthermore, these metrics provide valuable insights into model performance, offering a theoretical framework to explain why certain predictive models succeed or fail across different types of time series.

2 Methodology

We describe two metrics—Spectral Predictability and the largest Lyapunov Exponent—used to assess a time series' forecastability prior to model training. We provide detailed explanations of how each metric is computed and interpreted in the context of identifying intrinsic predictability in time series data.

2.1 Spectral Predictability

Spectral Predictability [7] quantifies the concentration and regularity of frequency components in a time series, serving as a proxy for its complexity in the Fourier domain. Time series with clear periodic patterns (e.g., seasonality) exhibit dominant frequency peaks, while highly irregular or noisy series have energy dispersed across a wide range of frequencies. In this context, predictability is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
KDD '25, Toronto, ON, Canada.
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1454-2/25/08
https://doi.org/10.1145/XXXXXX.XXXXXX

inversely related to the spectral entropy—a measure of disorder in the frequency domain.

Given a de-trended time series, $\boldsymbol{y} = (y_0, y_1, \ldots, y_{T-1})$, we first compute its power spectral density (PSD) using the Fast Fourier Transform [5]. Let p_i denote the normalized power of the i-th frequency component. The spectral entropy is given by:

$$H_a(\boldsymbol{y}) = \sum_i p_i \log_a p_i \tag{1}$$

The Spectral Predictability score can be defined as:

$$\Omega(\boldsymbol{y}) = 1 - \frac{H_a(\boldsymbol{y})}{\log_a(2\pi)}, \tag{2}$$

where a is the logarithmic base, typically set to e or 2. Normalizing by $\log_a(2\pi)$ bounds $\Omega(\boldsymbol{y})$ in $[0, 1]$, with higher values indicating lower spectral complexity and greater forecastability

The intuition behind this metric is that the complexity of a time series in the Fourier domain is directly related to its forecastability. For example, a flat spectrum indicates high unpredictability, as maximum spectral entropy corresponds to a uniform distribution of energy across all frequencies, where all possible frequencies contribute equally to the time series, making it highly complex and difficult for any model to forecast. Conversely, a constant time series exhibits zero spectral entropy and therefore has the highest spectral predictability.

To mitigate spectral leakage, we apply a Hann window before computing the Fourier transform [11, 13]. The metric can be computed globally or within a moving window to detect local changes in predictability. It is computationally efficient with time complexity $O(T \log T)$, making it practical for large-scale analysis.

2.2 Lyapunov Exponents

While spectral predictability captures harmonic structure in the frequency domain, it does not differentiate between deterministic chaos and stochastic noise. To address this, we complement it with Lyapunov Exponents [4], which measure the sensitivity of a dynamical system to initial conditions in the time domain. This metric provides insight into the system's stability and long-term behavior.

Given a time series $\boldsymbol{y} = (y_0, y_1, ..., y_{T-1})$, we first reconstruct its state space via time-delay embedding:

$$\mathbf{x}_t = \left(y_t, y_{t+\tau}, \ldots, y_{t+(m-1)\tau}\right) \in \mathbb{R}^m \tag{3}$$

where m is the embedding dimension and τ is the delay. Each vector $\mathbf{x}_t$ represents the system's state at time t in the reconstructed phase space.

To estimate the largest Lyapunov exponent, we track how the distance between initially close state vectors diverges over time. For each embedded state $\mathbf{x}_t$, we identify its nearest neighbor $\mathbf{x}'_t$, with initial separation:

$$\delta_0 = \|\mathbf{x}_t - \mathbf{x}'_t\| \tag{4}$$

We then observe how this separation evolves over a fixed number of time steps Δt:

$$\delta(\Delta t) = \|\mathbf{x}_{t+\Delta t} - \mathbf{x}_{t'+\Delta t}\| \tag{5}$$

The largest Lyapunov exponent is estimated as the average exponential rate of divergence:

$$\lambda = \frac{1}{\Delta t} \log \frac{\|\delta(\Delta t)\|}{\|\delta_0\|} \tag{6}$$

A positive λ indicates exponential divergence and chaotic behavior, implying reduced forecastability. A non-positive λ (zero or negative) suggests stability and higher forecastability.

In practice, we average λ over multiple state pairs to improve robustness. Since the estimation depends on accurate local trajectory tracking, it requires a sufficiently long and dense time series. Based on our experiments, we recommend using at least $100 \times m$ data points and limiting sparsity to below 0.7. This method is more computationally intensive than spectral analysis, typically with $O(T^2)$ complexity, but provides valuable insight into chaotic behaviors of a time series

Note that a dynamical system has a full spectrum of Lyapunov exponents $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, one per dimension in the reconstructed phase space. Each exponent λ_i measures the average exponential rate of divergence along a specific direction. In this work, we estimate only the *largest* Lyapunov exponent, defined as $\lambda = \max_i \lambda_i$, which dominates the system's long-term behavior.

3 Experiments

We conduct a series of experiments on both synthetic and real-world datasets to evaluate whether spectral predictability and Lyapunov exponents effectively reflect the intrinsic forecastability of time series. Our goals are twofold: (1) to validate that these metrics correlate with time series complexity and predictability, and (2) to demonstrate their alignment with downstream forecasting performance across different levels of data granularity.

3.1 Forecastability of a Synthetic Example

Experiment Setup. To illustrate the behavior of the two metrics, we construct a synthetic time series composed of five consecutive segments with increasing complexity and decreasing forecastability. These five segments of time series are shown in Figure 1, including a pure sine wave, a multi-frequency wave, a noisy multi-frequency wave, a Lorenz system trajectory, and white noise.

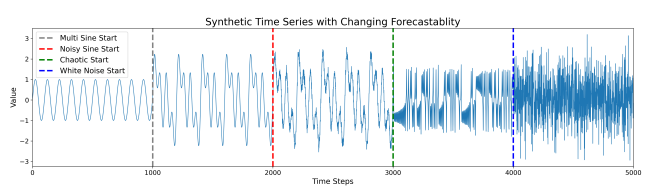

Figure 1: Five segments with increasing complexity and decreasing forecastability: a pure sine wave, a multi-frequency wave, a multi-frequency wave with additional random noise, a trajectory from Lorenz chaotic system and white noise.

Results. Figure 2 shows the moving spectral predictability with a window size of 200 (top) and moving largest Lyapunov Exponent with a window size of 300 (bottom) computed over the synthetic time series. We can see that both metrics respond consistently with our expectations: spectral predictability decreases, and the

Lyapunov exponent increases, as the underlying signal becomes more chaotic or noisy.

Spectral predictability fluctuates due to the use of a moving window, where each window may not contain full periodic cycles. However, the Fourier transform assumes that the input time series contains complete cycles. Incomplete cycles can introduce noisy spikes in the spectrum. In addition, the sudden increase in the Lyapunov Exponent plot or the sudden drop in Spectral Predictability before each segment occurs because they are computed in a moving window manner. When the window contains two different types of time series, it becomes much harder to forecast.

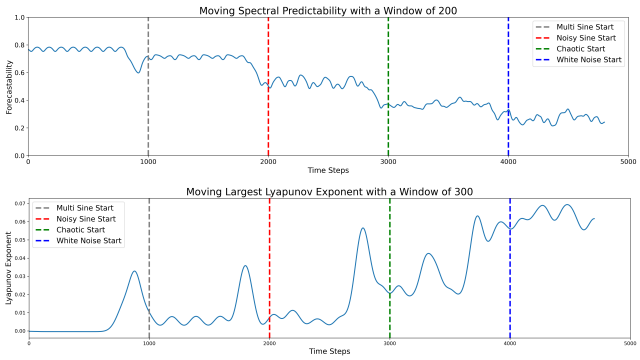

Figure 2: Top: moving spectral predictability with a window size of 200 over the synthetic time series. Bottom: moving largest Lyapunov Exponent with a window size of 300.

Note that the reason why the spectral predictability is not zero for the last white noise segment is that it reaches zero only when the spectrum follows a perfectly uniform distribution. This occurs only when the white noise time series is sufficiently long.

While neither metric can distinguish chaos from randomness directly, both serve as strong indicators of overall signal complexity and forecastability. This experiment also highlights their potential utility in identifying distributional shifts or regime changes in non-stationary time series.

3.2 Sensitivity Study of Metrics to Time Series Length and Sparsity

In this section, we evaluate how the two metrics respond to variations in time series length and sparsity—two key factors often encountered in real-world applications such as retail demand forecasting. Ideally, forecastability scores should decrease as sparsity increases, reflecting the loss of informative structure. We also examine sensitivity to time series length to ensure that, given sufficiently long sequences with similar characteristics, the metrics produce stable and consistent values across different lengths. This stability is important for enabling fair comparisons across different lengths.

Experiment Setup. We use the same five types of synthetic time series from the previous section, but vary their lengths (from 50 to 300) and sparsity rates (from 0% to 95%) by randomly zeroing out values. For each type of synthetic time series, we generate 100 sequences with different initial conditions and system parameters, varying both length and sparsity rate. We then compute spectral predictability and the largest Lyapunov exponent for each configuration. Figure 3 and Figure 4 shows how spectral predictability and Lyapunov Exponent change with series length (left) and sparsity rate (right). The shaded areas in the plots represent two standard deviations.

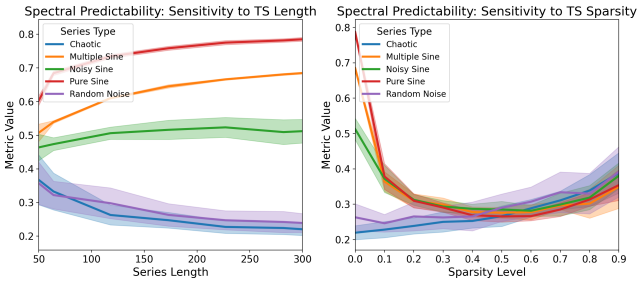

Figure 3: The sensitivity of the spectral predictability to varying time series length and sparsity. (Note: Higher spectral predictability indicates easier-to-forecast series.)

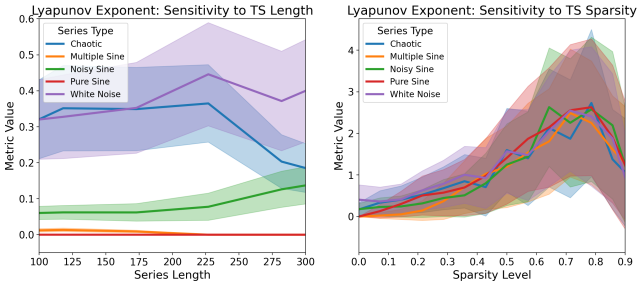

Figure 4: The sensitivity of the Lyapunov Exponent to varying time series length and sparsity. (Reminder: the lower the Lyapunov Exponent, the easier the time series is to forecast)

Results. We observe several patterns from Figure 3. For unpredictable series, longer sequences slightly reduce spectral predictability, while sparsity has a mild inflating effect. For moderately predictable series, predictability remains stable across lengths. For highly predictable series, longer lengths improve predictability, but increased sparsity sharply reduces it. We can conclude that spectral predictability takes sparsity into account and is not significantly affected by length, making it a comprehensive metric for determining forecastability

From Figure 4, we can observe that given sufficient length, the Lyapunov exponent isn't affected much by length. Additionally, increasing sparsity will increase the Lyapunov exponent, which is expected as the sparsity make time series harder to predict. But we can also see excessive sparsity (> 0.8) may decreases the Lyapunov exponent and falsely indicates the system as stable. With sufficient length and moderate sparsity, the Lyapunov exponent reliably captures forecastability. However, at extreme sparsity levels (>0.8), it may falsely indicate stability.

Spectral Predictability	Daily	Weekly	Lyapunov Exponents	Daily	Weekly
L0 (total)	0.374 ± 0.0	0.394 ± 0.0	L0 (total)	0.0 ± 0.0	0.081 ± 0.0
L1 (category)	0.358 ± 0.015	0.341 ± 0.008	L1 (category)	0.0 ± 0.0	0.055 ± 0.078
L2 (department)	0.339 ± 0.026	0.333 ± 0.047	L2 (department)	0.051 ± 0.125	0.169 ± 0.112
L3 (product)	0.246 ± 0.08	0.264 ± 0.057	L3 (Item)	0.833 ± 1.495	0.231 ± 0.982

Table 1: Spectral Predictability and Lyapunov Exponents across hierarchy levels

3.3 Forecastability vs. Prediction Errors on the M5 Dataset

To validate the practical utility of the two metrics, we study their correlation with actual prediction errors on a real-world dataset. We use the M5 forecasting competition dataset, which includes hierarchical sales time series across multiple levels—total sales, state, category, department, and product—at daily granularity.

Experiment Setup. We compute spectral predictability and Lyapunov exponents for each time series in the M5 dataset across multiple aggregation levels, including total, category, department, and product, as well as two temporal frequencies: daily and weekly. To assess their relationship with actual forecast accuracy, we train several forecasting models—including ETS [6], RecursiveTabular [16], and Chronos [1]—on each subset and compute the Weighted Absolute Percentage Error with help of AutoGluon library [16].

Results. Table 1 displays the Spectral Predictability and Lyapunov Exponents across multiple hierarchy levels and two temporal frequencies: daily and weekly. We compute the metrics for each time series at each level, and the reported standard deviations reflect the variance across different series within that level. For Spectral Predictability, higher values indicate greater predictability and thus higher forecastability. For the Lyapunov Exponent, a value of zero indicates a stable system, while positive values suggest chaotic behavior and lower forecastability. From the table, we observe that for daily time series, Level 0 (total daily unit sales) exhibits the highest forecastability. Forecastability generally decreases as we move to lower levels in the hierarchy, for both daily and weekly time series. Additionally, for Level 3 (product-level) series, aggregating the data from daily to weekly significantly improves forecastability, as reflected by both metrics.

To investigate whether these observations correlate with actual model prediction errors, we trained multiple models—including ETS, RecursiveTab, and Chronos—on various subsets of the M5 dataset across different hierarchy levels and temporal frequencies. We then plotted their prediction errors against the pre-computed spectral predictability scores, as shown in Figure 5.

Based on Figure 5 and Table 1, we observe that spectral predictability generally shows a negative correlation with WAPE, suggesting that higher predictability is associated with better model performance. In contrast, Lyapunov exponents exhibit a positive correlation with WAPE, indicating that more chaotic time series are more difficult to forecast. These correlations are strong for both daily and weekly frequencies (with $r =\sim 0.9$).

This experiment, together with the study on the synthetic time series dataset, demonstrates that both forecastability metrics reliably capture the inherent predictability of time series and show

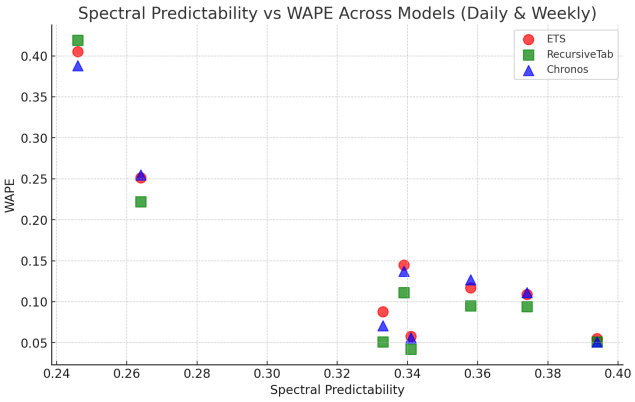

Figure 5: Spectral predictability vs prediction WAPE across multiple levels and time frequencies in M5

strong correlation with downstream forecasting performance. However, we note that this analysis does not provide sufficient evidence to determine which forecasting model performs best under varying levels of forecastability.

4 Discussion

We propose using two metrics—Spectral Predictability and the largest Lyapunov Exponent—to assess the forecastability of time series prior to model training. Our goal is to provide a lightweight, model-agnostic way to evaluate whether a time series is inherently predictable, enabling practitioners to focus resources on tractable forecasting tasks and adopt alternative strategies for more chaotic or noisy series.

Our experiments demonstrate that both metrics capture key aspects of time series complexity and correlate strongly with downstream forecasting performance. Spectral Predictability, based on frequency-domain entropy, offers a robust and computationally efficient measure that performs well across a wide range of data conditions. It is particularly useful when time series are short or moderately sparse. Lyapunov Exponents, which quantify sensitivity to initial conditions in the reconstructed phase space, provide complementary insights into the stability of dynamical behavior, but require longer sequences and lower sparsity to be reliable.

As a practical guideline, we find that spectral predictability scores below 0.2 or Lyapunov exponents above 1.0 are indicative of low forecastability. Spectral Predictability is stable with as few as 100 time steps, whereas Lyapunov estimation typically requires at least $100 \times m$ time steps, where m is the embedding dimension. Comparing spectral predictability to that of white noise with matched length

and sparsity further improves interpretability. Practitioners should exercise caution when interpreting results for highly sparse or short time series, such as monthly or yearly retail data, where metric stability may degrade.

Beyond guiding modeling strategy, these metrics have broader utility. Persistently low or unstable values can signal data quality issues such as insufficient history or structural noise. When computed over sliding windows, they can serve as indicators of distributional shift or regime change, helping determine when to retrain forecasting models. In this way, they also support model interpretability and monitoring.

Future work includes integrating these metrics into automated model selection pipelines and using them to inform active learning, anomaly detection, and dynamic retraining in real-time forecasting systems.

References

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. 2024. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815* (2024).

[2] Yossi Aviv. 2003. A time-series framework for supply-chain inventory management. *Operations Research* 51, 2 (2003), 210–227.

[3] Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, et al. 2022. Deep learning for time series forecasting: Tutorial and literature survey. *Comput. Surveys* 55, 6 (2022), 1–36.

[4] Jonathan B Dingwell. 2006. Lyapunov exponents. *Wiley encyclopedia of biomedical engineering* (2006).

[5] Pierre Duhamel and Martin Vetterli. 1990. Fast Fourier transforms: a tutorial review and a state of the art. *Signal processing* 19, 4 (1990), 259–299.

[6] Everette S Gardner Jr. 1985. Exponential smoothing: The state of the art. *Journal of forecasting* 4, 1 (1985), 1–28.

[7] Georg Goerg. 2013. Forecastable component analysis. In *International conference on machine learning*. PMLR, 64–72.

[8] James D Hamilton. 2020. *Time series analysis*. Princeton university press.

[9] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. 2024. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. 6555–6565.

[10] Bryan Lim and Stefan Zohren. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A* 379, 2194 (2021), 20200209.

[11] Douglas A Lyon. 2009. The discrete fourier transform, part 4: spectral leakage. *Journal of object technology* 8, 7 (2009).

[12] John T Mentzer, William DeWitt, James S Keebler, Soonhong Min, Nancy W Nix, Carlo D Smith, and Zach G Zacharia. 2001. Defining supply chain management. *Journal of Business logistics* 22, 2 (2001), 1–25.

[13] Nicolas Pielawski and Carolina Wählby. 2020. Introducing Hann windows for reducing edge-effects in patch-based image segmentation. *PloS one* 15, 3 (2020), e0229839.

[14] Damien Power. 2005. Supply chain management integration and implementation: a literature review. *Supply chain management: an International journal* 10, 4 (2005), 252–263.

[15] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. 2018. Deep state space models for time series forecasting. *Advances in neural information processing systems* 31 (2018).

[16] Oleksandr Shchur, Ali Caner Turkmen, Nick Erickson, Huibin Shen, Alexander Shirkov, Tony Hu, and Bernie Wang. 2023. AutoGluon–TimeSeries: AutoML for probabilistic time series forecasting. In *International Conference on Automated Machine Learning*. PMLR, 9–1.

[17] Rui Wang, Yihe Dong, Sercan O Arik, and Rose Yu. 2023. Koopman Neural Operator Forecaster for Time-series with Temporal Distributional Shifts. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=kUmdmHxK5N