EMPIRICALLY ASSESSING THE PREDICITIVTY OF TOURISM DEMAND DATA

RAUAN KYRYKBAYEV, GANG LI, AND CHRIS ZHANG

Abstract. Accurate forecasting of tourism demand is critical for policy and business planning, yet remains challenging due to the inherent complexity and vulnerability of tourism demand to external shocks. This study introduces a novel predictivity metric based on Weighted Permutation Entropy (WPE) for assessing the intrinsic predictivity of tourism demand data. Building on the limitations of existing entropy measures, particularly Sample Entropy (SampEn) and Multiscale SampEn, WPE is proposed for its effectiveness in capturing both ordinal and amplitude dynamics of the tourism demand, especially under external shocks, such as the COVID-19 pandemic. Using monthly tourist arrival data from Australia, the study assess the predictivity of tourism demand across different temporal scales and lengths. The findings demonstrate that WPE-based predictivity metric is effective during both stable and volatile periods, which is empirically validated by a strong inverse relationship between the predictivity metric and forecasting error. The study provides actionable insights for enhancing tourism demand forecasting by optimizing data aggregation scales and adapting the predictivity metric during volatile periods.

Contents

1. Introduction

The tourism industry is a vital driver of economic growth for many nations. Accurate tourism demand forecasting is essential in providing critical insights into emerging trends, evolving tourist preferences, and market dynamics, enabling policymakers and businesses to make informed decisions [12]. The increasing availability of data and advanced forecasting techniques has fueled efforts to refine these techniques and push the boundaries of predictive potential in tourism data.

However, despite these advancements, accurately predicting tourism demand remains a complex challenge [28]. Tourism demand is influenced by a multitude of factors, including seasonality, economic cycles, geopolitical events, and unexpected disruptions such as pandemics or natural disasters, all of which introduce significant variability and uncertainty into international tourist flows [21]. These complexities limit the effectiveness of traditional forecasting models, highlighting the need for a deeper understanding of the *intrinsic predictivity* of tourism demand data - the theoretical upper bound of forecasting accuracy determined by the inherent structure of the data itself [14]. In the context of tourism, *intrinsic predictivity* refers to the theoretical maximum accuracy future tourist arrivals can be achieved from historical data patterns [28].

This is particularly relevant, as a recent data-driven meta-method, the RobustSTL-ARIMA-LSTM model proposed from the work of Li et al. [12], has demonstrated high forecasting accuracy (e.g., up to 98 % for Hong Kong tourism data). These results raise fundamental questions about the true predictivity limits of tourism demand data and the reliability of existing predictivity metrics. If forecasting models achieve high accuracy in some cases but fail in others, understanding which predictivity metrics provide the most robust insights into *intrinsic predictivity* of tourism demand data is essential for future research and practical applications.

Several studies have proposed methods for evaluating predictivity, predominantly based on *Information Theory* and employing entropy measures [28]. For instance, Song et al. [20] used variations of *Shannon entropy* to calculate the maximum predictivity of human mobility, and Pennekamp et al. [14] suggested to quantify *intrinsic predictivity* of ecological time series with *Weighted Permutation Entropy*. Despite the importance of predictivity in tourism demand forecasting, limited research has investigated this area in depth. Zhang et al. [28] assessed the predictivity of tourist arrival data of Hong Kong by applying *Sample Entropy* (`SampEn`) and *Multiscale SampEn* (`MulSampEn`) to quantify the complexity and derive maximum predictivity using Fano's inequality. However, there is one critical research gap within this study: these entropy measures are not capable of accurately quantifying the complexity of non-stationary time series [18].

The limitation of these measures lies in their methodology, particularly the tolerance distance r, which is defined as a proportion of the time series' standard deviation. As a result, time series affected by external shocks exhibit inflated standard deviations, leading to a larger r value [18]. This broader threshold increases the likelihood of detecting more matching patterns [18], resulting in lower entropy and misleading high predictivity estimates, even though the actual predictivity of the series has decreased. This limitation raises concerns about the reliability of predictivity assessments on time series with significant disruptions.

Addressing this challenge is the main focus of this research and is essential for developing more robust and reliable metric for assessing the predictivity of tourism demand, especially in the face of unforeseen disruptions [16]. This study proposes a novel *Weighted Permutation Entropy* (`WPE`)-based metric for assessing the predictivity of tourism demand data. The motivation for using `WPE` lies in its sensitivity to temporal dynamics and structural complexity within time series, which other entropy measures often fail to capture [7]. Thus, this work aims to answer the following research question: **How to define a metric that effectively captures the intrinsic predictivity of tourism demand data, particularly under conditions of volatility and disruption?**

Building on the work of Zhang et al. [28] and other related studies, this study assesses the predictivity of tourism demand data and offers the following key contributions.

(1) The study proposes a novel and robust metric based on `WPE` for assessing the predictivity of tourism demand, which is effective in capturing the impacts of external shocks, such as the *COVID-19* pandemic.
(2) The research identifies the effect of data characteristics, specifically the length and temporal scale of time series data, on predictivity. This helps ensure that forecasting models are built using data with characteristics that offer the highest theoretical potential for accurate prediction.
(3) By understanding the relationship between predictivity and actual forecasting performance, practitioners can better select forecasting models that fully leverage the predictive potential of data.

These contributions are particularly valuable for risk mitigation strategies within the tourism sectors. For instance, understanding how predictivity declines during external shocks and how quickly it recovers provides an evaluation of the reliability of forecasting models. As a result, government and businesses can formulate more resilient and adaptive strategic plans, minimizing potential losses.

The remainder of this paper is structured as follows: Section 2 evaluates potential metrics to assess the predictivity of tourism demand data. Section 3 presents the calculation and parameter tuning of the predictivity metric based on Weighted Permutation Entropy (WPE). Section 4 empirically evaluates and compares predictivity metrics on Australia's tourist arrival statistics, across time series before and after the *COVID-19* pandemic, and explores the relationship between forecasting performance, data characteristics, and these predictivity metrics. Section 6 discusses the implications and presents the conclusion of the study.

2. Literature Review

This section reviews general predictivity measures from the perspective of time series complexity, which are rooted in two main theoretical frameworks: the *Dynamical Systems Theory* and *Information Theory* [27]. Subsequently, it will focus on entropy measures as a means to quantify *intrinsic predictivity*, discussing their theoretical foundation and applicability to tourism demand forecasting. Finally, it will examine existing applications of these measures in tourism and highlight the research gap this study aims to address.

2.1. General Predictivity Measures.

Predictivity in time series forecasting can be categorized into two main types: *realized predictivity* and *intrinsic predictivity*. Among them, *realized predictivity* refers to the actual accuracy achieved by a specific forecasting model when applied to a dataset. For example, the *Root Mean Squared*

Error (`RMSE`) of a time series forecast quantifies its *realized predictivity*. In contrast, *intrinsic predictivity* represents the theoretical upper limit on how well any model can forecast a time series based on the inherent deterministic and stochastic components in the data [14]. The *realized predictivity* of forecasting models can then be used to indicate how effectively these models capture the underlying *intrinsic predictivity*.

This study focuses on assessing the *intrinsic predictivity* of tourism demand, particularly in univariate time series. While multivariate models often offer improved forecasting potential, they involve complex predictor selection issues that are beyond the scope of this work [10].

Entropy measures offer a practical approximation of *intrinsic predictivity* by capturing the underlying structure and randomness of time series data. In general, time series that are highly regular (e.g., periodic) are more predictable, while those that are chaotic or purely stochastic exhibit reduced predictivity [5]. Measures of complexity aim to quantify this spectrum and distinguish between deterministic and random behavior [3].

Yao et al. [27] identified two foundational approaches to quantifying complexity:

- the dynamical systems approach, which relies on system trajectories and initial conditions; and
- the information theory approach, which assesses the randomness and structural order in data.

These frameworks form the basis for entropy measures of predictivity.

2.1.1. *Dynamical Systems Perspective.* From the dynamical systems perspective, predictivity is linked to the rate of error growth and the amount of information produced by the system over time [5]. The key measures in this context are the *Lyapunov exponent* (LE) that quantifies the time interval T_p on which the system is predictable [5], and the *Kolmogorov-Sinai* (KS) entropy, which is a measure of the rate of information loss due to unpredictability [2]. However, Boffetta et al. [5] noted several limitations of LE in real-world systems, particularly those with finite time or noisy observations. These challenges have motivated the use of modified methods such as the *finite-size Lyapunov exponent* (FSLE) for empirical applications [5].

Kolmogorov-Sinai entropy, while theoretically sound, is difficult to compute for empirical datasets, especially those with small data volumes or structural breaks, which is common in tourism demand data. These limitations have prompted researchers to exploit the statistical properties of the data generated by the system to study its dynamical behaviour using entropy measures derived from information theory [4].

2.1.2. *Information Theory Perspective.* In information theory, predictivity is inversely related to randomness. A time series with a well-defined structure exhibits lower entropy and therefore higher predictivity, while a fully random sequence has maximum entropy and is inherently unpredictable [2]. Entropy quantifies the average uncertainty in a sequence of outcomes and serves as a foundation for several complexity measures [2].

The most widely known entropy measures for time series include *Shannon Entropy*, *Permutation Entropy* (PE), *Sample Entropy* (SampEn), and their variants. These measures differ in the way they encode information, handle noise, and account for data attributes [2, 3] that are especially relevant in tourism demand data.

Shannon Entropy: Introduced by Claude Shannon in 1948, *Shannon Entropy* provides a foundational measure of uncertainty by calculating the probability distribution of time series values [2]. While it is simple to compute, it assumes stationarity and does not account for temporal or ordinal relationships, making it less suitable for non-stationary tourism data characterized by seasonality and structural shifts.

Permutation Entropy (PE): PE extends *Shannon Entropy* by incorporating ordinal structure and was proposed by Bandt and Pompe [3]. It analyzes the frequency of different order patterns within a time series and is robust to noise [3]. However, PE ignores amplitude information and may produce misleading results in time series with low resolution [2], which are typical challenges in tourism datasets.

To overcome this, several *PE* variants have been developed. For example, *Weighted Permutation Entropy* (`WPE`) incorporates amplitude information by weighting patterns based on local variability [7]. Other enhancements include *Multiscale WPE* (MSWPE) [25] and *Fine-Grained PE* (FGPE) [26], which capture dynamics at different scale or levels of detail.

Sample Entropy (SampEn): `SampEn` assesses the regularity of patterns by computing the probability that similar sequences remain similar when extended [17]. It is less sensitive to short time series than *Approximate Entropy* (AppEn) but relies heavily on the tolerance parameter r, which is typically set as a fraction of the standard deviation [2]. In volatile conditions, such as during a pandemic, this leads to inflated r values and unreliable predictivity assessments.

Given these properties, entropy measures have been increasingly applied to domains such as tourism, where complex and non-linear dynamics prevail [11].

2.2. **Entropy Measures for Tourism Demand Predictivity.** The application of entropy measures to tourism demand forecasting requires careful consideration of the data's specific characteristics, successful applications in related fields provide a basis for their use. For example, Song et al. [20] effectively employed entropy to quantify predictability in human mobility, a domain closely related to tourism. This demonstrates the potential of entropy to capture underlying regularities and constraints that influence demand patterns, justifying further exploration of entropy measures in the context of tourism demand.

The suitability of entropy measures to tourism demand data depends on their ability to handle its complex characteristics: strong seasonality, non-stationary behavior, and sensitivity to external shocks [21]. These challenges complicate the selection of a suitable entropy measure.

Sample Entropy has been used in tourism demand predictivity study due to its noise resistance and effectiveness in identifying recurring patterns [28]. However, its reliance on the tolerance parameter r makes it unreliable in the presence of structural breaks, such as those induced by the *COVID-19* pandemic.

Shannon Entropy and *PE*provide simpler alternatives but fail to account for magnitude fluctuations or higher-order dynamics [2]. In contrast, `WPE` offers a balanced approach by capturing both ordinal and amplitude information, making it more suitable for tourism demand data that demonstrate sudden volatility or long-term structural changes. Table 1 provides a comparison of the entropy measures discussed, highlighting their key features, strengths, limitations, and applicability to tourism demand data.

TABLE 1. Summary of Shannon Entropy, Permutation Entropy, WPE and SampEn.

Basis of entropy measure	**Strengths**	**Weaknesses**	**Applicability to Tourism Demand**	**Practical Challenges**
		Shannon Entropy		
Probabilistic distribution of data values	Simple to compute and interpret	Sensitive to noise and outliers; ignores temporal structure	Applicable to stationary demand	Requires sufficient data to estimate reliable probability distributions
		Permutation Entropy (PE)		
Ordinal patterns: analyzes relative ordering of data points within sequences	Robust to noise, applicable to deterministic and stochastic systems	Ignores amplitude variations and limited to ordinal patterns	Suitable for detecting structural changes and seasonal patterns in tourism trends	Choice of the embedding dimension can affect results
		Weighted Permutation Entropy (`WPE`)		
Ordinal patterns with weighting function to capture local variations	Improved PE with the weighting function to capture the magnitude of fluctuations	Weighting function design can introduce bias, requires parameters tuning and sufficient data length for seasonal time series	Evaluating structural and magnitude-based predictivity trends	Sensitive to the choice of the parameters (m, τ)
		Sample Entropy (SampEn)		
Pattern regularity: evaluates similarity between subsequences within a tolerance (r)	Ignores minor noise fluctuations; uses relative matching between patterns	Unreliable in the presence of structural breaks due to inflated tolerance parameter (r)	Useful for assessing regularities in tourism demand	Requires adoption of a resampling approach for nonstationary data

Zhang et al. [28] provided a methodological framework for assessing the maximum predictivity of tourism data using entropy measures. Their approach involved calculating `SampEn`(`MulSampEn`) and then relating these entropy values to maximum predictivity using Fano's inequality. This approach demonstrates that entropy measures can be adapted to quantify the theoretical limits in the predictivity of tourism demand data.

The methodological framework proposed by [28], while offering valuable insights towards quantifying the predictivity of tourism demand using entropy measures, exhibits several limitations that require further investigation. First, their study lacks a robust rationale for the selection of `SampEn` as a suitable measure for assessing the *intrinsic predictivity* of tourism demand data. This lack of clear justification raises concerns about the fundamental validity of using `SampEn` to accurately reflect the inherent characteristics of tourism demand data. Second, the scope of their analysis regarding external shocks is limited to the *SARS* outbreak in 2003. While significant, the impact of *SARS* was relatively contained and short-lived compared to the profound and prolonged impact caused by the *COVID-19* pandemic [22]. This limited consideration of external shocks raises questions about the generalizability and robustness of their findings, particularly in the context of more severe and prolonged global crises that significantly alter tourism demand patterns.

2.3. **Summary.** Although entropy measures have shown promise in evaluating the predictivity of time series, their application to tourism demand forecasting remains limited and fragmented. Most existing studies emphasize *realized predictivity* through model-dependent evaluations [28], with insufficient attention to intrinsic data characteristics.

The work Zhang et al. [28] has made notable progress using `SampEn` (`MulSampEn`) to quantify the *intrinsic predictivity* of tourism demand data. However, their approach overlooks a key methodological limitation inherent to this entropy measure, which leads to incorrect evaluation of predictivity under external shocks, such as *COVID-19* pandemic.

This limitation arises from the use of the tolerance parameter r, which is defined relative to the standard deviation (σ) of the time series. During periods of large-scale fluctuations, this dependence causes (r) to increase, making more embedding vectors appear similar when compared using Chebyshev distance [18]. Consequently, the computed entropy artificially decreases, resulting in higher predictivity when, in fact, the underlying dynamics have become more uncertain. To overcome this issue, the present study proposes `WPE` as a more robust alternative for capturing the *intrinsic predictivity* of tourism demand data.

3. The Predictivity of Tourism Demand Data

The research gap outlined previously motivates the present study to propose *Weighted Permutation Entropy* (`WPE`) as a superior alternative. `WPE`'s enhanced ability to capture temporal dynamics, coupled with its robustness and efficiency, offers a more reliable and practical approach to assessing the *intrinsic predictivity* of tourism demand, particularly in the case of disruptive events where existing measures fall short. Additionally, its simplified parameter tuning and lower computational complexity compared to other entropy measures enhance its practicality for real-world applications [2]. This section presents the calculation and parameter tuning of `WPE` for assessing the *intrinsic predictivity* of tourism demand data.

3.1. **Formalization of Tourism Demand Forecasting.** In this study, tourism demand forecasting task is defined as predicting the number of future tourist arrivals at destinations using past tourism data. The forecasting relies exclusively on historical data presented as a univariate time series [28].

Let the time series be represented by the vector $Y^T = \{y^{(1)}, y^{(2)}, .., y^{(T)}\}$ with length T, where $\{y^{(i)}\}_{i=1}^{k}$ serves as the input, and $\{y^{(i)}\}_{i=k+1}^{k+\delta}$ represents the forecasted values, with δ indicating the number of steps ahead to predict [28].

The forecasting model, denoted as F, uses the input segment $\{y^{(i)}\}_{i=1}^{k}$ to estimate the future values $\{y^{(i)}\}_{i=k+1}^{k+\delta}$ and the accuracy of the model is assessed by metrics such as the *Root Mean Square Error* (`RMSE`) [28].

3.2. **Predictivity Based on Weighted Permutation Entropy (WPE).** The `WPE` extends the standard *Permutation Entropy* (PE) by capturing the dynamic changes in the amplitude of the time series [14]. The embedding procedure forms $T-\tau(m-1)$ vectors $Y(1), Y(2), ..., Y(T-\tau(m-1)), Y(i) = [y(i), y(i+1), ..., y(i+\tau(m-1))], i = 1, ..., T-\tau(m-1)$. Then the m real values of a vector *Y(i)* are sorted in increasing order: $\{y(i+k_1-1) \leq y(i+k_2-1) \leq ... \leq y(i+k_m-1), 1 \leq k_1, k_2, ..., k_m \leq m\}$. As a result, any vector *Y(i)* is mapped into a vector $K_m = [k_1, k_2, ..., k_m]$, where the original value is replaced by its ordinal index, which is represented as $\{0, 1, ..., m-1\}$. Each k_m's contribution to the probability mass function is weighted by its variance $w(K_m) \equiv var(Y_i)$ and the weighted probability of each permutation π is estimated by:

$$p_w(\pi) = \frac{\sum\limits_{t \leq T-m} w(K_m) \cdot \delta(\phi(K_m), \pi)}{\sum\limits_{t \leq T-m} w(K_m)} \tag{1}$$

where $\delta(k_i, k_j) = 1$ if $k_i = k_j$ and $\delta(k_i, k_j) = 0$ otherwise [14]. The weighted permutation entropy of order $m \geq 2$ is defined as:

$$h_w(m) = -\sum_{\pi \in K_m} p_w(\pi) \log_2 (p_w(\pi)) \tag{2}$$

Then the `WPE` is normalized by the possible maximum PE, and the predictivity (`WPE`) is calculated as:

$$\Psi_{h_w(m)} = 1 - \frac{h_w(m)}{h_{\max}(m)} \tag{3}$$

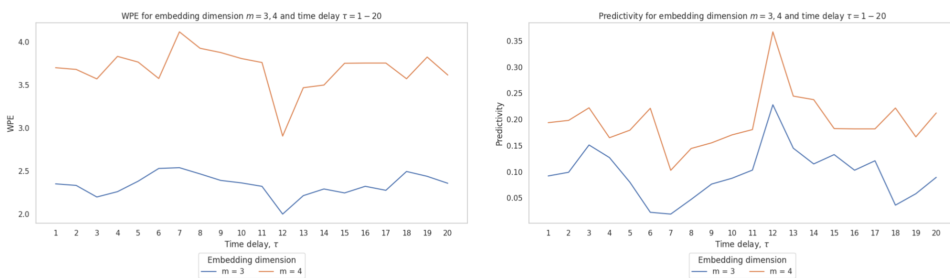
where $h(m)_{max} = \log_2(m!)$ and $0 \leq h_w(m) \leq h(m)_{max}$ [14]. Thus, the predictivity value is within the range $[0, 1]$, where 0 represents a completely random and unpredictable time series, and 1 indicates a fully deterministic and perfectly predictable time series. This approach of estimating predictivity is conceptually different from [28], as it links predictivity relative to the bound of complete randomness or unpredictability in a time series.

3.2.1. *Parameter tuning for WPE.* The calculation of `WPE` relies on the selection of key parameters: the embedding dimension (m) and the time delay (τ). These parameters significantly influence the resulting entropy value and, consequently, the predictivity metric.

Choosing appropriate values for m and τ is essential for obtaining meaningful results. Riedl et al. [19] provides some heuristics, but there is no single rule of thumb that fits all cases. The optimal parameter selection depends on the characteristics of the time series being analyzed, such as its length, sampling frequency, and underlying dynamics [19].

For measuring the complexity of the main oscillation (seasonal cycle), Riedl et al. [19] suggests setting the time delay τ equal to the period length of the main oscillation, for instance, $\tau = 12$ for monthly data with annual seasonality. To measure the processes that possess an inherent cycle represented by triggering events Riedl et al. [19] recommends $\tau = 1$. As for the embedding dimension m, it is recommended to choose the maximum according to $N > 5m!$, where N is the length of the time series.

Figure 1 illustrates an example of `WPE` parameters tuning for a monthly time series of length $N = 406$. The optimal delay parameter within the range 1-20 is observed at $\tau = 12$, corresponding to a local minimum in `WPE` [19]. Given the length of the time series, an embedding dimension of $m = 3$ is appropriate. As the embedding dimension should reveal forbidden permutations and ensure reliable statistics, it typically requires an average of 100 counts per permutation [8]. Therefore, for this monthly time series, the optimal delay parameter $\tau = 12$ and the embedding dimension $m = 3$. Conversely, Figure 1b demonstrates that the predictivity(`WPE`) increases with a larger embedding dimension, as an insufficient number of permutations is observed at $m = 4$.



(A) WPE

(B) Predictivity(WPE)

FIGURE 1. The figure (A) illustrates the `WPE` for embedding dimensions $m = 3, 4$ and time delay $\tau = 1 - 20$, and (B) the predictivity of a time series. The main oscillation for monthly data is identified at $\tau = 12$, where the `WPE` is minimal and predictivity is maximum. This indicates that a time delay of 12 months provides the most accurate assessment of predictivity for monthly time series.

4. The Empirical Validation of The Predictivity Metric

This section empirically assesses the predictivity of tourism demand data by addressing the core research question: **How to define a metric that effectively captures the intrinsic predictivity of tourism demand data, particularly under conditions of volatility and disruption?**

The section begins by presenting the validation framework and methods used to validate the metric. This is followed by a detailed rationale and description of the data used in the analysis. The predictivity of Australia's tourism demand is then assessed using the proposed predictivity metric and compared against the baseline metric based on `SampEn` (`MulSampEn`). Subsequently, the empirical validation of the proposed predictivity metric is carried out within the established framework. All experiments are implemented using the `Python` programming platform.

4.1. **The predictivity metric validation framework.** The framework, illustrated in Figure 2, assesses the effectiveness of the predictivity metric by defining its correlation with the *realized predictivity* of tourism demand data.

The validation framework involves the following steps: (1) Calculate the predictivity for a diverse set of tourism demand datasets. (2) Apply established time series forecasting models (e.g., `ARIMA`, `ARIMAX`) to these datasets and quantify their *realized predictivity* with forecasting error (e.g. `RMSPE`). (3) Analyze the correlation between the predictivity metric and forecasting error.

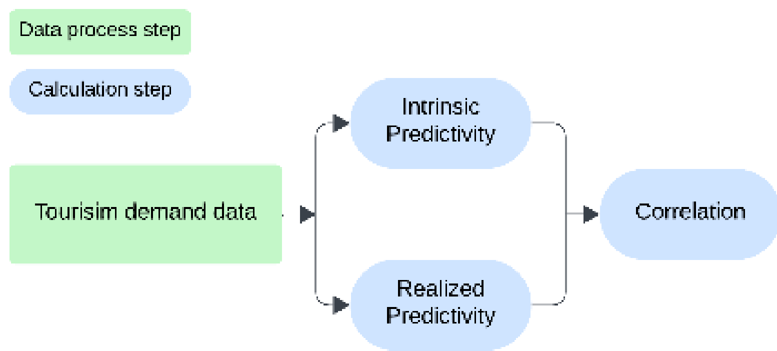


FIGURE 2. The predictivity metric validation framework.

A strong negative correlation would indicate that higher predictivity (lower complexity) corresponds to lower forecasting errors (higher *realized predictivity*), thus validating the proposed metric.

4.1.1. *Methods used in the predictivity validation framework.* The methods used in the predictivity metric validation framework include the predictivity metrics, the forecasting models, the forecasting error, and the correlation coefficients.

An *Autoregressive Integrated Moving Average* (`ARIMA`) and an (`ARIMAX`) models are applied to predict tourist arrivals, as `ARIMA` is one of the most commonly used benchmark models in tourism demand forecasting [9]. The *realized predictivity* of a time series is quantified using *Root Mean Squared Percentage Error* (`RMSPE`), which measures predictive accuracy by comparing predicted and observed values. Lower values indicate that better *realized predictivity* is achieved.

$$\text{RMSPE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{|y_t - y'_t|}{y_t}\right)^2} \tag{4}$$

where T is the total number of observations, y'_t is predicted tourist arrivals and y_t actual tourist arrivals.

The rank-based non-parametric correlation coefficients *Kendall*'s and *Spearman*'s correlation coefficients are employed to assess the relationship between the *intrinsic* and *realized predictivity.* Non-parametric methods are suitable when the underlying data distribution is unknown or when dealing with small sample sizes [6], conditions that are common for tourism demand data. These coefficients assess the strength and direction of the monotonic association between predictivity metric rankings and the corresponding `RMSPE` rankings.

Kendall's coefficient is a conjoint unweighted rank measure, which reflects the agreement between the rankings and provides a direct probabilistic interpretation that i, j are ranked in the same order in both rankings [24]. The formula for *Kendall*'s coefficient is listed as below:

$$\tau = \frac{C - D}{\binom{n}{2}} \tag{5}$$

where C is the number of concordant pairs, D is the number of discordant pairs,

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

the number of possible pairs for n datasets.

Spearman's correlation coefficient measures the strength and direction of a monotonic relationship between two ranked variables, it is computed by first determining the differences between the ranks of the corresponding values and then applying the formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{6}$$

where d_i represents the difference between ranks, and n is the total number of observations [13].

Both correlation coefficients range from -1 to 1. If a coefficient value is negative, this implies that the predictivity values are inversely correlated with `RMSPE`. A value of -1 indicates a perfect negative monotonic relationship, where higher predictivity values correspond to lower `RMSPE` values consistently across all datasets. In this context, a negative correlation coefficient suggests that as predictivity increases (or complexity decreases), the `RMSPE` tends to decrease, which is a desirable outcome indicating stronger potential of the predictivity metric. A positive correlation value, or close to zero, on the other hand, implies little to no meaningful relationship between the predictivity metric and forecasting error. This would suggest that the metric fails to capture the *intrinsic predictivity* of the data.

4.2. **Rationale and description of the Australian tourism demand data.** This study uses monthly statistics of short-term international visitor arrivals to Australia, sourced from the *Australian Bureau of Statistics* [1]. There are three key reasons for selecting Australian tourism demand data for this study: First, tourism has been one of Australia's fastest growing industries and is a major export industry for Australia, losing nearly A$9 billion every month during the *COVID-19* pandemic and facing job losses of over 300 000 people [15]. This makes Australia an economically significant case study due to the important role tourism plays in its economy. Second, this study aims to assess the effectiveness of the predictivity metric, `WPE`, by applying it to multiple datasets, ensuring its robustness. Australia is well suited for this purpose, as it provides long-term and state-level tourism data, allowing a wide range of time series to be tested at different temporal and spatial scales. Moreover, no previous study has specifically assessed the predictivity of Australia's tourism demand, making this research the first to explore this. Therefore, the predictivity assessment of Australian tourism demand can provide a deeper understanding of a significant and diverse tourism market, supporting the development of more accurate forecasting models.

The data consist of 406 monthly records of international tourist arrivals in the Australian states and territories from January 1991 to October 2024, and include nine distinct datasets that are presented in Figure 3. It is observed that the tourist flow increased from 1991 to February 2020, followed by a sharp decline due to the *COVID-19* travel restrictions introduced in March 2020. The international border restrictions were subsequently lifted on July 6, 2022 [23].

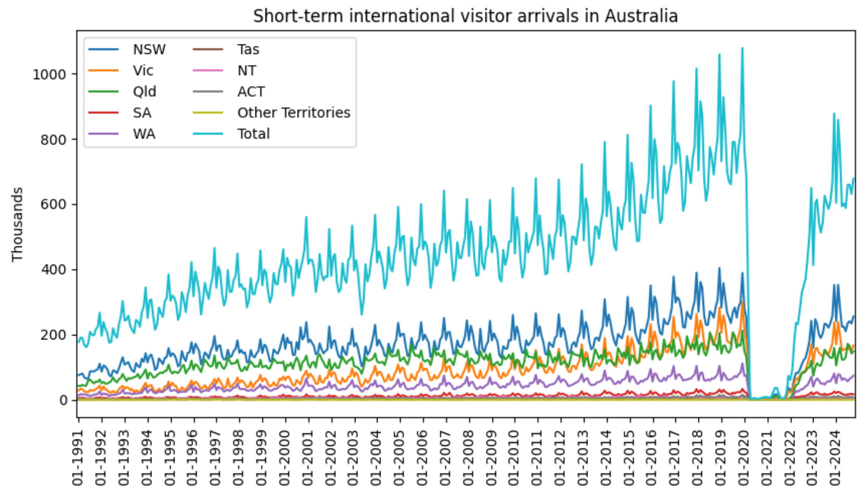


FIGURE 3. The graph shows monthly international tourist arrivals in Australia and its states and territories from January 1991 to October 2024.

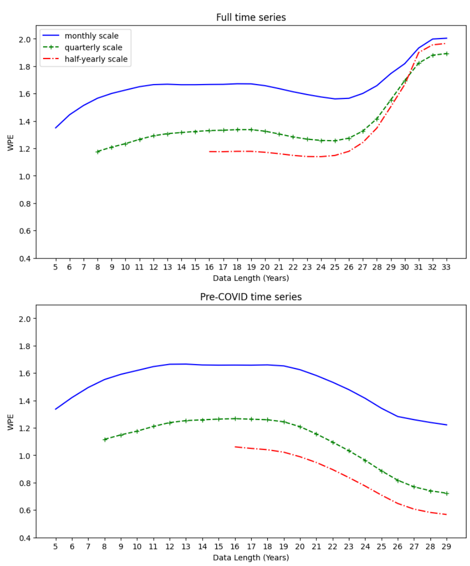
All analyses are conducted using the complete historical datasets, each dataset is divided into sub-series of varying lengths using a sliding-window approach. This approach involves sequentially moving a sub-series of a specific length along the full time series to generate multiple overlapping sub-series. During the validation with forecasting models, each sub-series is split into training and test sets. Forecasting models are trained on the training sets, and forecasts are generated for the test periods. Forecast accuracy is then assessed by calculating the forecasting error over these test sets. The analysis is conducted across three periods: the full period (January 1991 to October 2024), the pre-COVID period (up to February 2020), and the post-COVID period (from February 2020 onward). The monthly time series are also scaled to generate corresponding quarterly (3-month) and half-yearly (6-month) series to define the impact of data characteristics on predictivity.

4.3. **Predictivity and complexity assessment using WPE and SampEn (MulSampEn).** This section compares *intrinsic predictivity* and complexity measured by `WPE` and `SampEn(MulSampEn)` across time series of varying lengths and scales (granularity).

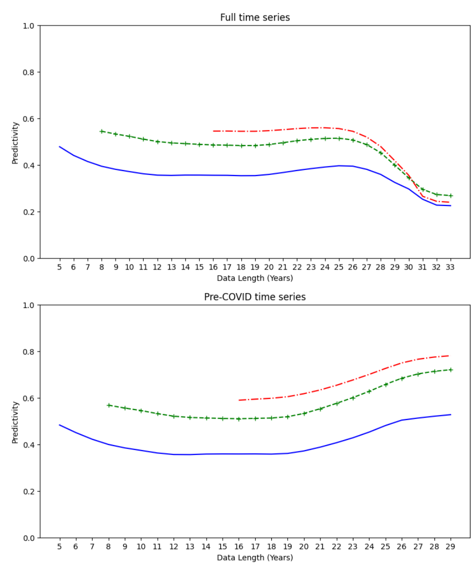
The assessment is conducted on on the time series of the "Total" dataset, with data lengths ranging from 5 to 32 years and temporal scales of monthly, quarterly and half-yearly, separately on the full and the pre-COVID periods. Mean predictivity and entropy values are calculated using a sliding-window approach: for each sub-series, the entropy measures and predictivity metrics are calculated, and the average value is taken across all sub-series of a given length.

4.3.1. *Predictivity assessment with `WPE`.* The complexity of the time series is measured using `WPE`, with the embedding dimension set to $m = 3$ and the delay parameter set to $\tau = 12$ for monthly, 4 for quarterly, and 2 for half-yearly series. The corresponding predictivity metric is calculated using eq. (3). According to the guideline by [19], the minimum required sub-series length must be not less than $5m!$ to ensure reliable entropy estimation. Thus, the minimum lengths for quarterly and half-yearly time series are 8 and 16 years, respectively.

Figure 4 presents the complexity and predictivity assessed with `WPE` for the full and pre-COVID time series. As observed for the pre-COVID series, where external shocks are absent, the complexity tends to decrease, and predictivity (`WPE`) improves with larger scale, being maximized at the largest scale (half-yearly) and the longest series (29 years). However, the impact of COVID disrupts this trend in the full series, leading to an increase in complexity and a decrease in predictivity. This suggests that COVID increases volatility, leading to reduced predictivity, which in turn makes forecasting more challenging to model.



(A) Mean `WPE`.



(B) Mean Predictivity(`WPE`).

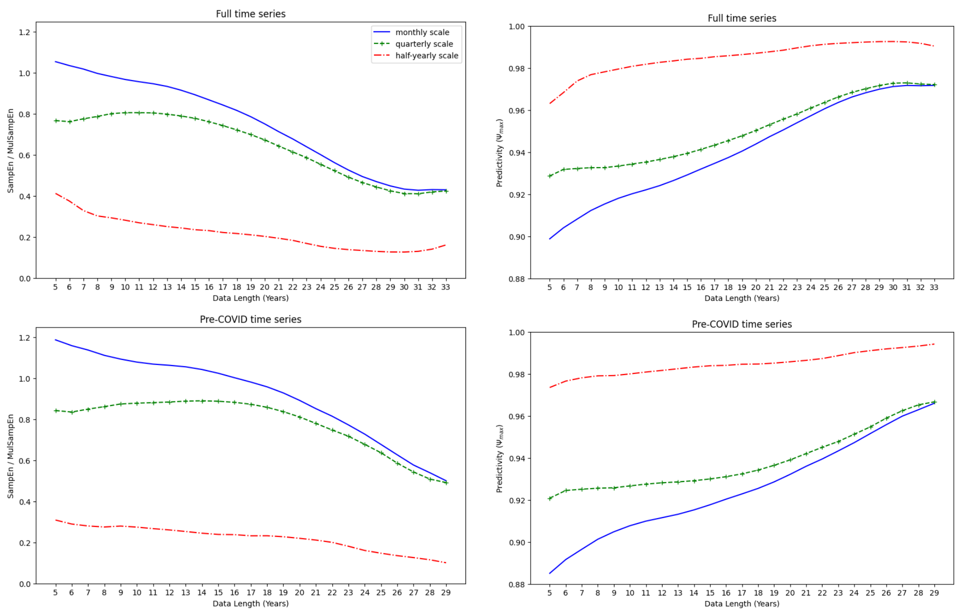
FIGURE 4. The `WPE` and predictivity of the full and pre-COVID series by time scales: monthly, quarterly and half-yearly. The COVID period introduces additional volatility, resulting in a decrease in predictivity, which makes predictions less reliable and more complex to model. This requires adjustments to forecasting models to account for structural breaks and changing dynamics.

The results indicate that larger data scales can enhance predictivity during stable periods, whereas significant disruptions can alter the underlying structure and predictivity of the data. This demonstrates that `WPE` effectively captures changes in predictivity caused by external shocks.

4.3.2. *Predictivity assessment with `SampEn`, `MulSampEn`.*
The complexity of the monthly time series is assessed using `SampEn`, and the quarterly and half-yearly series are assessed using `MulSampEn`. Similar to `SampEn`, `MulSampEn` captures the complexity of the time series data with multiple scaling levels [28].

The resulting predictivity metric based on these entropy measures, denoted as $\Psi_{\max}$, is calculated following the approach outlined by [28].

For SampEn and MulSampEn the following parameters are set: the embedding dimension $m = 3$, delay $\tau = 1$, and tolerance distance $r = 0.65 \times \sigma$. The SampEn, MulSampEn, and Ψ_{max} of the full and pre-COVID time series, as shown in Figure 5, exhibit almost no difference in the trend between the two series, suggesting that the effects of COVID are not strongly reflected by these entropy measures. This limitation arises because the standard deviation (σ) increases when there is a significant change in the trend, which expands the tolerance distance r. As a result, a larger number of embedding vectors, compared by *Chebyshev* distance, fall within r and are counted as similar, leading to a decrease in complexity and increase in Ψ_{max}. This effect persists even when the coefficient value of r is reduced, exposing a methodological limitation in SampEn, MulSampEn.



(A) Mean SampEn, MulSampEn (B) Mean Ψ_{max}

FIGURE 5. The SampEn, MulSampEn and Ψ_{max} of the full and pre-COVID series by time scales: monthly, quarterly and half-yearly. SampEn, MulSampEn is unable to detect disruptions like COVID in tourism demand data, due to a methodological limitation rooted in the tolerance parameter r.

The limitation of SampEn and MulSampEn highlights the potential of WPE in assessing the *intrinsic predictivity* of tourism demand data, especially under external shocks.

4.3.3. *Predictivity assessment over external shock.* The *COVID-19* pandemic caused global shifts in all industries, including tourism. Understanding how this structural shock affected the predictivity of tourism data is crucial for tourism analysts in developing robust forecasting models that incorporate structural shifts and external disruptions. This section examines how the impact of *COVID-19* on predictivity varies across different data lengths. These insights are valuable for practitioners aiming to identify the optimal data length for forecasting models.

As shown in 6, the impact of *COVID-19* is evident in the widening gap between the mean predictivity (WPE) of the full and pre-COVID time series. This widening occurs due to the growing proportion of COVID-affected sub-series as the data length increases.

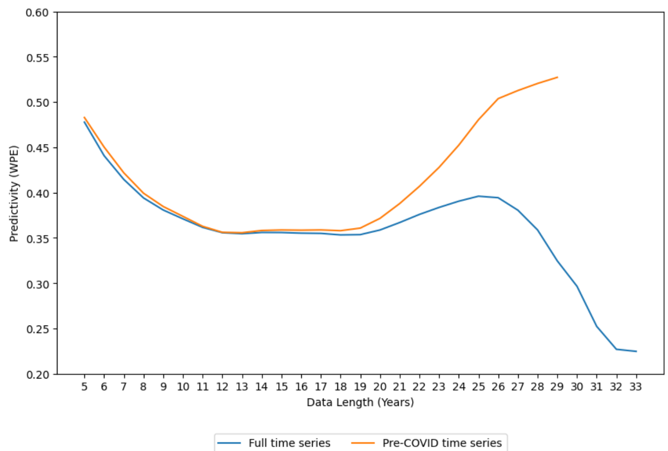
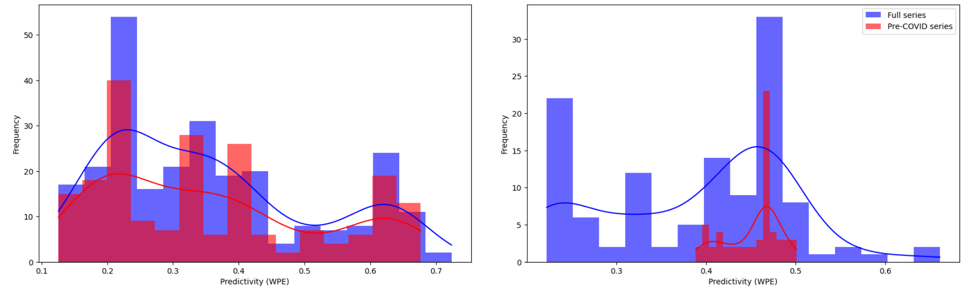


FIGURE 6. Mean Predictivity(WPE) of the full and pre-COVID monthly time series.

To illustrate this effect, the predictivity distributions are compared for two different data lengths: 144 months (12 years), where the gap is minimal, and 288 months (24 years), where the gap becomes more pronounced.

In Figure 7a, the predictivity distributions of the full and pre-COVID series are nearly identical for a data length of 144 months (12 years), as the proportion of COVID sub-series is low and has an insignificant impact on the mean predictivity. However, for a data length of 288 months (24 years) Figure 7b, the proportion of COVID-affected subseries with low predictivity, ranging between 0.0 and 0.3, becomes substantial, leading to a decrease in mean predictivity (WPE). It is also observed that some sub-series within the COVID period exhibit higher predictivity values compared to those from the pre-COVID period. To further investigate the effect of *COVID-19*, the trend of sub-series predictivity for the two data lengths is illustrated in Figure 8.

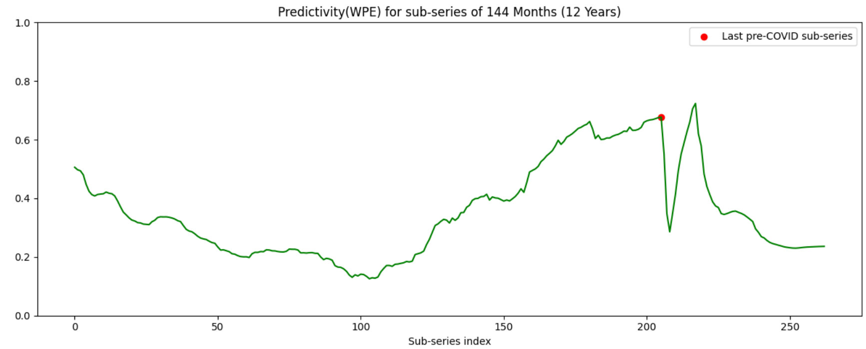


(A) Distribution of Predictivity (`WPE`) for sub-series of 144 Months (12 Years).

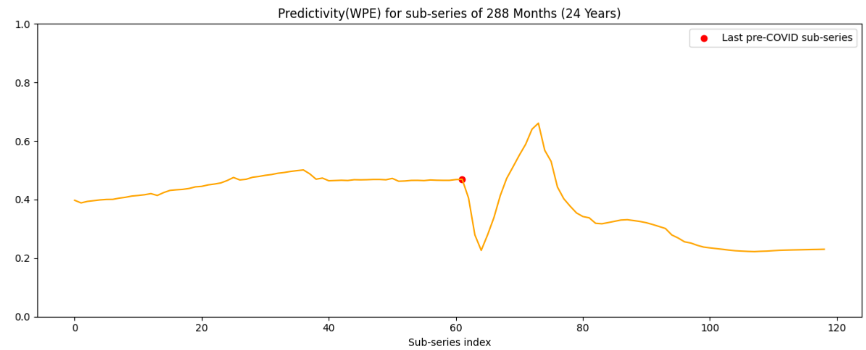
(B) Distribution of Predictivity (`WPE`) for sub-series of 288 Months (24 Years).

FIGURE 7. The predictivity (`WPE`) distribution of sub-series with lengths of 144 months (12 years) and 288 months (24 years) for the full and pre-COVID series. The predictivity distribution of the pre-COVID series becomes narrower with increasing series length, as more recurring patterns are captured. Therefore, longer time series can enhance the reliability of forecasting models in the absence of external shocks. In contrast, the COVID-affected sub-series cause a broader distribution in longer series, reflecting increased volatility and reduced pattern consistency.

The red dot in Figure 8 represents the predictivity of the last pre-COVID sub-series. The effect of *COVID-19* is observed within the length of the oscillation period, or the delay parameter ($\tau = 12$), of `WPE` for monthly time series. During the initial three months, COVID has a negative impact on predictivity, followed by a positive effect over the subsequent nine months, reaching the maximum point. These effects, occurring within the oscillation period, are due to the increased weight (standard deviation) of similar patterns that include the *COVID-19* months. This shows a limitation of `WPE`: it can be influenced by short-term high-volatility events if those events create new, regularly repeating patterns, such as the year-on-year drop in tourist arrivals caused by *COVID-19*. As a result, it may hide a longer-term decrease in actual predictivity caused by increasing randomness in the data. This decrease becomes evident after the peak point, when the impact of *COVID-19* leads to a steady decline in predictivity (`WPE`), caused by the emergence of increasingly varied and irregular patterns in the data. Therefore, to accurately assess the predictivity of time series impacted by external shocks that alter their cyclical patterns, the delay parameter τ needs to be adjusted.



(A) Trend of Predictivity (WPE) for sub-series of 144 Months (12 Years).



(B) Trend of Predictivity (WPE) for sub-series of 288 Months (24 Years).

FIGURE 8. The predictivity (WPE) trend of sub-series with lengths of 144 months (12 years) and 288 months (24 years) for the full series. The effect of *COVID-19* on the predictivity is evident with the first sub-series including the initial month affected by the pandemic. As the pandemic disrupted the cyclical pattern of the time series, the assessment of predictivity (WPE) on the COVID-affected periods must be conducted with an adjusted delay parameter τ.

As Figure 8 demonstrates, during the pre-COVID period, the volatility in predictivity of the shorter series (12 years) is considerably higher compared to the longer series (24 years). As the data length increases, the range of the predictivity (WPE) distribution of the pre-COVID series narrows, indicating that longer time series better capture the underlying structure and repeating patterns. The longer 24-year series also better illustrates the extent of *COVID-19*'s impact on predictivity: while the pre-COVID predictivity was around 0.4, the stabilized post-COVID predictivity levels, observed in sub-series with an index of 100 or higher, drop to approximately 0.2. Therefore, for cyclical data like tourism demand, longer time series offer more stable predictivity assessments than shorter ones. This suggests that tourism forecasters can benefit from using longer historical data, as it helps capture stable and repeatable trends, leading to more reliable predictions.

4.4. **The validation of the predictivity metric.** This section validates and compares the predictivity metrics to identify which of the two most effectively captures the *intrinsic predictivity* of the time series. An ARIMA model is applied to the pre-COVID time series. ARIMA consists of three key parameters: p, which represents the order of autoregression, q denotes the order of the moving average, and d indicates the degree of differencing applied to achieve stationarity. To determine the optimal model parameters, a grid search is performed over the ranges $p =$[0,1,2], $q =$[0,1,2], $d =$[0,1,2]. The best combination is then selected based on the lowest Akaike Information Criterion (AIC) [28].

The monthly pre-COVID time series of lengths 12 and 24 years are used to generate a 12-month forecast. This forecast horizon is selected for its alignment with the oscillation (main) period. For each destination dataset, the mean `RMSPE`, Ψ_{max} and Predictivity(`WPE`) are calculated to validate the predictivity metric with actual forecasting performance.

Table 2 summarizes the mean `RMSPE`, Ψ_{max}, and Predictivity(`WPE`) across different destinations and data lengths. Table 3 shows the correlation between these metrics and `RMSPE`.

TABLE 2. Mean `RMSPE`, Ψ_{max}, and Predictivity(`WPE`) by Destination and Data Length.

Destination	**Data Length 12 Years**			**Data Length 24 Years**		
	RMSPE	Ψ_{max}	**Predictivity(`WPE`)**	**RMSPE**	Ψ_{max}	**Predictivity(`WPE`)**
Total	0.1477	0.9101	0.3376	0.1506	0.9452	0.4490
NSW	0.1764	0.9139	0.2175	0.1680	0.9375	0.2785
Vic	0.2029	0.9226	0.5475	0.1805	0.9682	0.6496
Qld	0.1345	0.8970	0.1610	0.1323	0.9282	0.2142
SA	0.2477	0.9068	0.2120	0.2374	0.9412	0.2451
WA	0.2294	0.9255	0.2696	0.2079	0.9497	0.3019
Tas	0.5429	0.9508	0.1265	0.3982	0.9547	0.1340
NT	0.3489	0.9212	0.1121	0.4107	0.9429	0.0984
ACT	0.2355	0.8920	0.1751	0.2533	0.9394	0.1802

TABLE 3. Correlation results between `RMSPE` and Predictivity (WPE), and between `RMSPE` and Ψ_{max} across 12- and 24-year pre-COVID time series.

	Kendall's τ	**Kendall's p-value**	**Spearman's ρ**	**Spearman's p-value**
RMSPE vs Predictivity (WPE)				
12 Years	-0.3889	0.1802	-0.5333	0.1392
24 Years	-0.5556	0.0446	-0.6667	0.0499
RMSPE vs Ψ_{max}				
12 Years	0.3889	0.1802	0.4000	0.2861
24 Years	0.2222	0.4767	0.3167	0.4064

There is a negative correlation between predictivity(`WPE`) and `RMSPE`, which is statistically significant at 5 % level for the 24-year time series. This validates that lower forecasting error is associated with higher predictivity, indicating that more structured and less random time series tend to yield more accurate forecasts. The stronger negative correlation for the 24-year time series algins with the earlier findings, confirming that longer time series offer more consistent predictivity, as reflected in better aligned forecasting performance. In contrast, the positive correlation observed for Ψ_{max} highlights its limited effectiveness in assessing the *intrinsic predictivity* of tourism data.

4.4.1. *Predictivity validation over varied data scale.* This section validates the predictivity(`WPE`) by examining its relationship with forecasting error across different scales of temporal aggregation, thereby assessing how data scale influences predictivity in relation to underlying data characteristics.

Table 4 and Table 5 present the mean `RMSPE` of `ARIMA` models and predictivity(`WPE`) for the pre-COVID series of length 12 and 24 years. The results clearly indicate that the data scale has a significant impact on forecasting performance. The time series aggregated at quarterly and half-yearly scales exhibit higher predictivity, resulting in improved forecasting accuracy compared to monthly data. This is supported by a strong negative relationship between `RMSPE` and predictivity at larger data scales, as indicated by both Kendall's and Spearman's correlation coefficients in Table 6. These findings suggest that, in the absence of external shocks, tourism analysts can signficantly improve forecasting performance by using more coarsely scaled time series.

Table 4. Mean Predictivity(`WPE`) and `RMSPE` by Destination and Scale Level. Data length 12 years.

Destination	Mean RMSPE			Mean Predictivity		
	Monthly	**Quarterly**	**Half-Yearly**	**Monthly**	**Quarterly**	**Half-Yearly**
Total	0.1477	0.0546	0.0446	0.3376	0.5031	0.5419
NSW	0.1764	0.0687	0.0509	0.2175	0.3752	0.3858
Vic	0.2029	0.0675	0.0478	0.5475	0.6751	0.7242
Qld	0.1345	0.0646	0.0622	0.1610	0.2979	0.3920
SA	0.2477	0.0894	0.0640	0.2120	0.3810	0.4889
WA	0.2294	0.0571	0.0448	0.2696	0.4654	0.6628
Tas	0.5429	0.1459	0.0970	0.1265	0.2320	0.3388
NT	0.3489	0.1906	0.1533	0.1121	0.1875	0.1943
ACT	0.2355	0.1160	0.0862	0.1751	0.2781	0.4009

Table 5. Mean Predictivity(`WPE`) and `RMSPE` by Destination and Scale Level. Data length 24 years.

Destination	Mean RMSPE			Mean Predictivity		
	Monthly	**Quarterly**	**Half-Yearly**	**Monthly**	**Quarterly**	**Half-Yearly**
Total	0.1506	0.0310	0.0283	0.4490	0.6500	0.6647
NSW	0.1680	0.0401	0.0347	0.2785	0.4862	0.4824
Vic	0.1805	0.0466	0.0324	0.6496	0.8186	0.8405
Qld	0.1323	0.0382	0.0371	0.2142	0.3908	0.4722
SA	0.2374	0.0674	0.0571	0.2451	0.4107	0.4894
WA	0.2079	0.0369	0.0450	0.3019	0.5215	0.7411
Tas	0.3982	0.1786	0.1096	0.1340	0.3528	0.4261
NT	0.4107	0.1691	0.1652	0.0984	0.1424	0.1563
ACT	0.2533	0.1119	0.1128	0.1802	0.3182	0.4184

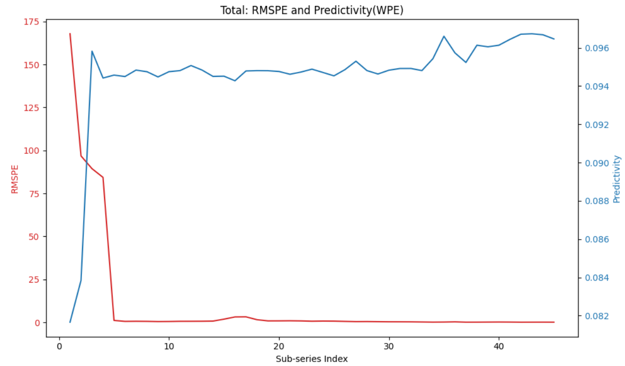
Table 6. Correlation results between Predictivity(`WPE`) and `RMSPE` across different data scales for pre-COVID time series of 12 and 24 years in length.

Data scale	**Kendall's τ**	**Kendall's p-value**	**Spearman's ρ**	**Spearman's p-value**
12-year time series				
Monthly	-0.3889	0.1802	-0.5333	0.1392
Quarterly	-0.6667	0.0127	-0.8000	0.0096
Half-yearly	-0.5556	0.0446	-0.7833	0.0125
24-year time series				
Monthly	-0.5556	0.0446	-0.6667	0.0499
Quarterly	-0.5556	0.0446	-0.7167	0.0298
Half-yearly	-0.6667	0.0127	-0.7833	0.0125

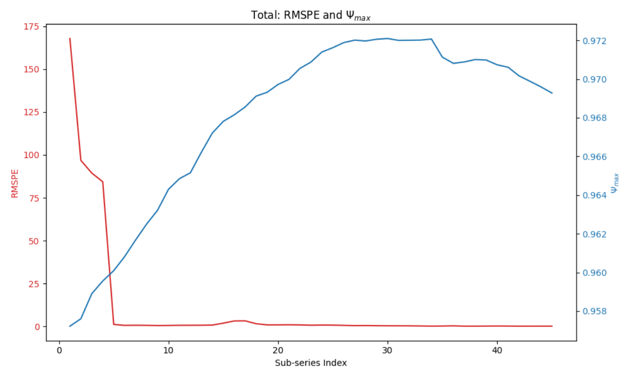
4.4.2. *Predictivity validation over external shock.* In this section the predictivity metric is validated on the COVID-affected time series. An `ARIMAX` model is employed for the COVID-affected series to account for external shocks. In addition to the `ARIMA` model parameters, `ARIMAX` incorporates an exogenous dummy variable to account for the influence of an external factor, the *COVID-19* pandemic, on the forecasting process, allowing for a more robust prediction under structural shifts.

Given the diverse impact of *COVID-19* across Australian states, the "Total" level is selected to evaluate the overall effect of the pandemic on the predictivity. To ensure an accurate assessment, the delay parameter τ is adjusted from 12 to 1, as the pandemic is considered as an external shock or triggering event that disrupted the original cyclical pattern.

Figure 9 illustrates the trend in the predictivity metrics and `RMSPE` across 45 sub-series of length 288 months (24 years). The first sub-series spans from March 1996 to February 2020, with the latter marking the beginning of the COVID-affected period. The results show that forecasting performance, indicated by a decrease in `RMSPE`, improves in parallel with increases in predictivity (`WPE`). In contrast, Ψ_{max} increases only gradually and exhibits weaker alignment with changes in `RMSPE`, suggesting that it is less responsive to structural variations in the data. This limited responsiveness reflects the inherent constraints of `SampEn` to short-term disruptions and evolving temporal dynamics, reinforcing the conclusion that `WPE` provides a more effective and adaptive metric for assessing the predictivity of tourism demand data with rapidly changing structures.



(A) `RMSPE` and Predictivity(`WPE`)



(B) `RMSPE` and Ψ_{max}

FIGURE 9. `RMSPE` (red line) and predictivity metrics (blue line), for 45 monthly sub-series of the "Total" dataset, each length of 24 years. The plot demonstrates that the model's adjustment to the *COVID-19* impact corresponds with an initial increase in the predictivity(`WPE`), followed by stabilization in both predictivity and forecasting performance. This pattern suggests that, after a short-term adaptation period, the forecasting model becomes more resilient to pandemic-related volatility. Such insights can be utilized to inform the selection of training windows that account for external shocks, enhancing model robustness and accuracy during periods of structural change.

TABLE 7. Correlation results between `RMSPE` and the predictivity metrics.

`RMSPE`	**Kendall's τ**	**Kendall's p-value**	**Spearman's ρ**	**Spearman's p-value**
Predictivity(`WPE`)	-0.5879	0.0000	-0.7611	0.0000
Ψ_{max}	-0.3636	0.0004	-0.5646	0.0001

The stronger negative correlation between `RMSPE` and predictivity (`WPE`), as shown in Table 7, further supports the superiority of `WPE` in capturing the *intrinsic predictivity* of tourism demand data. Overall, the results confirm that the metric based on `WPE` serves as a robust and reliable measure for assessing the predictivity of tourism demand data across both stable and volatile periods.

5. Discussion and implications

The empirical findings of this study offer profound insights into the assessment of tourism demand predictivity, particularly under volatile conditions induced by external shocks. This work demonstrates the effectiveness of `WPE` in capturing the *intrinsic predictivity* of tourism demand data and validates its superior robustness compared to `SampEn (MulSampEn)`.

5.1. Theoretical implications.

This research significantly advances the theoretical understanding of time series predictivity, especially within complex, non-stationary systems like tourism demand. From a theoretical standpoint, this study contributes to the broader literature on time series predictivity by proposing `WPE` as a more robust alternative for estimating the *intrinsic predictivity* of tourism demand data. This advancement is significant because it challenges the existing approach, where entropy measures are applied without sufficient consideration of their limitations.

By exposing a key flaw in `SampEn (MulSampEn)`, its reliance on the tolerance parameter (r), which inflates with increased standard deviation during structural breaks and leads to artificially high predictivity estimates, this work advances the Information Theory perspective on the robustness and reliability of entropy measures.

The research contributes to theoretical development by showing how *intrinsic predictivity* varies with data length and temporal scale. Theoretical frameworks in Information Theory and Dynamical Systems Theory assume that predictivity is tightly linked to structural order [5]. This study confirms that predictivity tends to stabilize and improve with longer time series and coarser aggregation (e.g., quarterly or half-yearly), thereby implying that longer and coarser data enhance the underlying structural order.

The study shows that the effectiveness of entropy measures depends on the system's context, thereby deepening our understanding of their appropriate application and reinforcing their theoretical importance as means for assessing *intrinsic predictivity.* This contributes to the theoretical role of entropy measures not only as a descriptor of complexity but also as a predictor of forecasting potential, making them much more useful for real-world planning and decision-making in complex environments.

5.2. Practical implications.

The practical implications of this study are particularly relevant for tourism analysts, policy-makers, and business strategists who rely on accurate and resilient forecasting to support planning, investment, and operational decisions. One of the most persistent challenges in tourism demand forecasting lies in determining the appropriate data length and temporal scale for model input [28], a decision that significantly affects predictive performance.

The study's findings provide concrete guidance on this issue. By empirically demonstrating that longer time series and coarser temporal scales (e.g., quarterly or half-yearly) yield higher *intrinsic predictivity*, the results suggest that practitioners can improve forecast accuracy by aggregating the data to a coarser scale. In practice, this insight helps streamline the forecasting pipeline by enabling more informed decisions about how to prepare input data to maximize predictive potential.

In volatile environments, such as during a global pandemic, the study demonstrates that `WPE` responds effectively to structural breaks. By adjusting `WPE` parameters, particularly the delay parameter, to reflect shifts in the data's underlying structure, practitioners can better capture the changes in the *intrinsic predictivity*. This is especially critical in real-world forecasting contexts, where abrupt changes in travel restrictions, health policies, or geopolitical events can drastically alter tourism demand patterns.

The validation of `WPE`-based predictivity metric against *realized predictivity* (`RMSPE`) confirms its effectiveness. Unlike `SampEn (MulSampEn)`, which may misleadingly indicate high predictivity during volatile periods, `WPE` provides a more truthful reflection of forecasting difficulty. This capability is invaluable for risk management and model selection. Forecasting models that are guided by `WPE`-based predictivity assessments are more likely to deliver robust performance, especially in high-uncertainty scenarios. Additionally, `WPE`'s computational efficiency and relatively simple parameter tuning, compared to other entropy measures, facilitate easier adoption by the industry practitioners.

Overall, this research introduces a superior metric for assessing the *intrinsic predictivity* of tourism demand. By integrating a theoretically grounded and empirically validated metric, this study not only advances the understanding of tourism demand predictivity but also equips practitioners with a more adaptive and robust tool for navigating uncertainty in a rapidly evolving global landscape.

6. Conclusion

Forecasting tourism demand plays a pivotal role in guiding strategic planning, resource allocation, and risk mitigation for governments and tourism-related businesses [12]. However, the volatility and complexity of tourism data, especially during external shocks like the *COVID-19* pandemic, make it difficult to assess the theoretical limits of forecasting accuracy. Accurately quantifying *intrinsic predictivity* is therefore essential to improve the reliability of forecasting models under both stable and disrupted conditions.

This research was motivated by the limitations of existing entropy measures, particularly `SampEn` and `MulSampEn`. There was a clear need for a more robust, interpretable, and sensitive metric capable of accurately assessing the *intrinsic predictivity* of tourism demand data, especially under conditions of volatility and sudden change.

The key contribution of this study is the introduction of a `WPE`-based predictivity metric for quantifying the *intrinsic predictivity* of tourism demand with greater reliability. Empirical results demonstrate that the proposed metric is more sensitive to structural changes and more strongly correlated with the *realized predictivity* compared to the predictivity metric based on `SampEn (MulSampEn)`. The study also reveals the impact of data characteristics, specifically verifying that coarsely scaled data increases predictive potential. These findings validate the effectiveness of `WPE` as a robust entropy measure for assessing the *intrinsic predictivity* of tourism demand data.

Despite these contributions, this study has several limitations:

(1) The analysis is restricted to the *COVID-19* pandemic as a representative external shock. While this event provides a relevant and impactful case study, broader generalizations to other types of disruptions require further investigation to assess the robustness of the proposed metric.

(2) The empirical analysis is based solely on tourism demand data from a single country (Australia). Although Australia's state level tourism demand presents a diverse context, the geographic specificity may limit the generalizability of the findings.
(3) The forecasting component focuses on univariate time series, potentially overlooking multivariate dependencies. Therefore, the maximum achievable forecasting accuracy may not have been achieved, and the relationship between *intrinsic* and *realized predictivity* may remain partially unverified.

Therefore, future research could extend the analysis to other types of disruptions to assess the generalizability of the predictivity metric (`WPE`) in capturing structural volatility beyond the *COVID-19* context. As well as its broader application to a wider range of tourism-related time series beyond international tourist arrivals. Expanding the study to include tourism demand data from multiple countries would enhance the external validity of the findings and uncover potential regional differences in predictivity behavior. Incorporating multivariate forecasting models that account for exogenous variables could offer a more complete understanding of the link between *intrinsic* and *realized predictivity*.

References

[1] Australian Bureau of Statistics ABS. 'table 11: Short-term movement, visitors arriving - intended state of stay: Original' [time series spreadsheet], 10 2024. URL https://www.abs.gov.au/statistics/industry/tourism-and-transport/overseas-arrivals-and-departures-australia/latest-release#data-downloads.

[2] A. Adelyanov, Evgenii Generalov, Wen Zhen, and Leonid Yakovenko. Using entropy in time series analysis. *Moscow University Physics Bulletin*, 79:415–425, 10 2024. doi: 10.3103/S0027134924700607.

[3] Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Physical review letters*, 88:174102, 05 2002. doi: 10.1103/PhysRevLett.88.174102.

[4] Chiara Barà, Riccardo Pernice, Cristina Angela Catania, Mirvana Hilal, Alberto Porta, Anne Humeau-Heurtier, and Luca Faes. Comparison of entropy rate measures for the evaluation of time series complexity: Simulations and application to heart rate and respiratory variability. *Biocybernetics and Biomedical Engineering*, 44(2):380–392, 2024. ISSN 0208-5216. doi: https://doi.org/10.1016/j.bbe.2024.04.004. URL https://www.sciencedirect.com/science/article/pii/S0208521624000287.

[5] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani. Predictability: a way to characterize complexity. *Physics Reports*, 356(6):367–474, 2002. ISSN 0370-1573. doi: https://doi.org/10.1016/S0370-1573(01)00025-4. URL https://www.sciencedirect.com/science/article/pii/S0370157301000254.

[6] Thierry Denœux, Marie-Hélène Masson, and Pierre-Alexandre Hébert. Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems*, 153(1):1–28, 2005. ISSN 0165-0114. doi: https://doi.org/10.1016/j.fss.2005.01.008. URL https://www.sciencedirect.com/science/article/pii/S0165011405000291.

[7] Bilal Fadlallah, Badong Chen, Andreas Keil, and José Príncipe. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E*, 87:022911, 2 2013. doi: 10.1103/PhysRevE.87.022911. URL https://link.aps.org/doi/10.1103/PhysRevE.87.022911.

[8] Joshua Garland, Ryan James, and Elizabeth Bradley. Model-free quantification of time-series predictability. *Phys. Rev. E*, 90:052910, 11 2014. doi: 10.1103/PhysRevE.90.052910. URL https://link.aps.org/doi/10.1103/PhysRevE.90.052910.

[9] Xiaoying Jiao, Gang Li, and Jason Li Chen. Forecasting international tourism demand: a local spatiotemporal model. *Annals of Tourism Research*, 83:102937, 2020. ISSN 0160-7383. doi: https://doi.org/10.1016/j.annals.2020.102937. URL https://www.sciencedirect.com/science/article/pii/S0160738320300815.

[10] Anton Kovantsev and Peter Gladilin. Analysis of multivariate time series predictability based on their features. pages 348–355, 11 2020. doi: 10.1109/ICDMW51313.2020.00055.

[11] Rob Law, Gang Li, Davis Fong, and Xin Han. Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, 75:410–423, 03 2019. doi: 10.1016/j.annals.2019.01.014.

[12] Xin Li, Xu Zhang, Chengyuan Zhang, and Shouyang Wang. Forecasting tourism demand with a novel robust decomposition and ensemble framework. 236:121388. ISSN 0957-4174. doi: 10.1016/j.eswa.2023.121388. URL https://www.sciencedirect.com/science/article/pii/S0957417423018900.

[13] Haoqian Pan and Yoshiyasu Takefuji. Enhancing heart disease feature analysis with spearman's correlation with p-values. *International Journal of Cardiology*, page 133207, 2025. ISSN 0167-5273. doi: https://doi.org/10.1016/j.ijcard.2025.133207. URL https://www.sciencedirect.com/science/article/pii/S0167527325002505.

[14] Frank Pennekamp, Alison C. Iles, Joshua Garland, Georgina Brennan, Ulrich Brose, Ursula Gaedke, Ute Jacob, Pavel Kratina, Blake Matthews, Stephan Munch, Mark Novak, Gian Marco Palamara, Björn C. Rall, Benjamin Rosenbaum, Andrea Tabi, Colette Ward, Richard Williams, Hao Ye, and Owen L. Petchey. The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89(2):e01359, 2019. ISSN 0012-9615. doi: 10.1002/ecm.1359.

[15] Tien Duc Pham, Larry Dwyer, Jen-Je Su, and Tramy Ngo. Covid-19 impacts of inbound tourism on australian economy. *Annals of Tourism Research*, 88:103179, 2021. ISSN 0160-7383. doi: https://doi.org/10.1016/j.annals.2021.103179. URL https://www.sciencedirect.com/science/article/pii/S0160738321000414.

[16] Efstathios Polyzos, Aristeidis Samitas, and Anastasia Spyridou. Tourism demand and the covid-19 pandemic: an lstm approach. *Tourism Recreation Research*, 46, 06 2020. doi: 10.1080/02508281.2020.1777053.

[17] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000. doi: 10.1152/ajpheart.2000.278.6.H2039.

[18] Joshua S. Richman, Douglas E. Lake, and J.Randall Moorman. Sample entropy. In *Numerical Computer Methods, Part E*, volume 384 of *Methods in Enzymology*, pages 172–184. Academic Press, 2004. doi: https://doi.org/10.1016/S0076-6879(04)84011-4. URL https://www.sciencedirect.com/science/article/pii/S0076687904840114.

[19] Maik Riedl, Andreas Oertel, and Niels Wessel. Practical considerations of permutation entropy: A tutorial review. *The European Physical Journal Special Topics*, 222, 06 2013. doi: 10.1140/epjst/e2013-01862-7.

[20] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010. doi: 10.1126/science.1177170. URL https://www.science.org/doi/abs/10.1126/science.1177170.

[21] Haiyan Song, Richard T.R. Qiu, and Jinah Park. A review of research on tourism demand forecasting: Launching the annals of tourism research curated collection on tourism demand forecasting. *Annals of Tourism Research*, 75:338–362, 2019. ISSN 0160-7383. doi: https://doi.org/10.1016/j.annals.2018.12.001. URL https://www.sciencedirect.com/science/article/pii/S0160738318301312.

[22] Haiyan Song, Gang Li, and Yanting Cai. Tourism forecasting competition in the time of covid-19: An assessment of ex ante forecasts. *Annals of Tourism Research*, 96:103445, 2022. ISSN 0160-7383. doi: https://doi.org/10.1016/j.annals.2022.103445. URL https://www.sciencedirect.com/science/article/pii/S0160738322000962.

[23] The Departmen of Home Affairs. All covid-19 border restrictions to be lifted. https://minister.homeaffairs.gov.au/ClareONeil/Pages/covid-border-restrictions-to-be-lifted.aspx, July 2022. Accessed: 2025-05-12.

[24] William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.*, 28:20, 11 2010. doi: 10.1145/1852102.1852106.

[25] Jianan Xia, Pengjian Shang, Jing Wang, and Wenbin Shi. Permutation and weighted-permutation entropy analysis for the complexity of nonlinear time series. *Communications in Nonlinear Science and Numerical Simulation*, 31(1):60–68, 2016. ISSN 1007-5704. doi: https://doi.org/10.1016/j.cnsns.2015.07.011. URL https://www.sciencedirect.com/science/article/pii/S1007570415002580.

[26] Paiheng Xu, Likang Yin, Zhongtao Yue, and Tao Zhou. On predictability of time series. *Physica A: Statistical Mechanics and its Applications*, 523:345–351, 2019. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2019.02.006. URL https://www.sciencedirect.com/science/article/pii/S0378437119301591.

[27] Weiguang Yao, Christopher Essex, Pei Yu, and Matt Davison. Measure of predictability. *Phys. Rev. E*, 69:066121, 06 2004. doi: 10.1103/PhysRevE.69.066121. URL https://link.aps.org/doi/10.1103/PhysRevE.69.066121.

[28] Yishuo Zhang, Gang Li, Birgit Muskat, Huy Quan Vu, and Rob Law. Predictivity of tourism demand data. *Annals of Tourism Research*, 89:103234, 2021. ISSN 0160-7383. doi: https://doi.org/10.1016/j.annals.2021.103234. URL https://www.sciencedirect.com/science/article/pii/S0160738321001122.

(A. 1) School of Information Technology, Deakin University, Melbourne, Burwood VIC 3125, Australia

Email address, A. 1: `rauan.kyrykbayev@gmail.com`

(A. 2) School of Information Technology, Deakin University, Melbourne, Burwood VIC 3125, Australia

Email address, A. 2: `gang.li@deakin.edu.au`

(A. 3) School of Information Technology, Deakin University, Melbourne, Burwood VIC 3125, Australia

Email address, A. 3: `chris.zhang@deakin.edu.au`