Forecasting campground demand in US national parks

Abstract:
Camping has grown from a recreational activity to an emerging tourism sector. In America’s national parks, this growth is amplified by increasing visitation and an occupancy limited by a mission to preserve the nation’s natural wonders. Forecasting future demand for campsites can not only aid administrators’ resource allocation, efficient management, and effective communication, but also provide valuable information to campers as they plan their vacations. This manuscript explores the unique nature of campground administration and tests a variety of forecasting methods to identify which best lends itself to the distinctive behavior of camping and the unique nature of campsites. An in-depth study of five popular campgrounds finds an ensemble model of neural network autoregression, k-nearest neighbors, exponential smoothing, and seasonal auto-regressive integrated moving average to be the most accurate prediction model.

Keywords:
national park; campground; demand forecasting; k-nearest neighbors; neural network autoregression; machine learning
Forecasting campground demand in US national parks

Camping, once novel recreation for the weary urban masses, is now a major form of tourism (Brooker & Joppe, 2014; Hogue, 2016). Within the United States, camping alone generates $166 billion annually (The Outdoor Industry Association, 2017). In 2015, Americans camped an average of 14.5 days per year (The Outdoor Foundation, 2017). However, compared to other sectors of the tourism industry, camping is relatively under-researched (Triantafillidou & Siomkos, 2013). A few limited studies on camping tourism have provided some insights into the motivations of campers, however. Viallon (2012) finds that Recreational Vehicle (RV) campers in Morocco are far more interested in the natural and social environments than the cultural environment surrounding the campgrounds. García-Pozo, Sánchez-Ollero, and Marchante-Lara (2011) examined the factors that determine campsite price through a hedonistic model, finding that location, the services available, and category of establishment (level of luxury) were among the most influential determinants for price.

However, research of campsites in national parks in the United States, which hosted over eight million camping overnight stays in 2017, remains understudied (National Park Service, 2018b). This deficit persists even as demand for national park campsites has reached a fever pitch. As a result, park managers may lack the understanding of what is driving demand, how to meet the increased demand, and how to predict future demand. Further complicating this problem is the unique mission of the National Park Service (NPS), which places preservation as the guiding mission of the agency. Therefore, increasing supply (e.g. building more campsites) is often not a viable management action. The upshot of this stalemate of supply and demand is a predicament for campground managers. Forecasting, based on historical reservation data, can provide a tool to manage tourism demand through inventory allocation strategy and dynamic pricing. Various forecasting methods have been
applied broadly in hotel demand forecasting, helping administrators improve revenue management and customer service (e.g. Weatherford & Kimes, 2003; Wu, Song, & Shen, 2017). However, to date, these tools have not been applied to the emerging sector of national park camping.

In this paper, we take six approaches to forecast campsite demand, five of which are already widely used to forecast tourism demand such as international tourist arrivals and hotel occupancy (Wu et al., 2017, Jiao & Chen, 2018): moving average (MA), exponential smoothing (ETS), seasonal auto-regressive integrated moving average (SARIMA), and neural network autoregression (NNAR). However, campsite demand manifest distinct characteristics from that of destinations and hotels since the campsites are often closed during certain seasons and booked far in advance due to lack of acceptable substitutes and limited access. This indicates a possibility that the existing methods for forecasting may not accurately capture the characteristics of campsite demand. Therefore, we introduce $k$-nearest neighbors (KNN) algorithm as an additional machine learning approach. Initial observation of the reservation data indicated similar seasonality, which fits KNN as the method to predict data points based on the most similar $k$ entries. We also develop an ensemble method that combines ETS, SARIMA, NNAR and KNN to achieve higher accuracy.

This study seeks to bridge the research gap both topic- and method-wise by analyzing demand for national park campgrounds and forecasting their future demand using a novel method on a decade of campsite reservation data from the Recreation Information Database (RIDB). RIDB is a service provided by the U.S. federal government containing data of historical reservations made online for campgrounds on federal lands (RIDB, n.d.). In doing so, we not only provide a new application and perspective on the depth of existing forecasting methodology, but also propose a new method. We begin by providing a review of the relevant literature and descriptive statistics of campsite demand at a park level, then apply an
array of forecasting methods to five individual campgrounds, and finally identify the best method for predicting the future demand of camping in America’s national parks.

**Literature Review**

In order to understand the approaches to the forecasting of national park campground demand, we first need to understand the possible drivers of demand, the nature of this tourism sector, the unique qualities of campers, and the array of forecasting methods at our disposal.

**Increasing Campsite Demand**

The dramatic increase in NPS campsite demand is inconsistent with the participation trend of camping overall. The Outdoor Foundation (2017) reports camping rates remaining relatively steady from 2013 to 2016 (the year of last available data) with 13.8% and 13.7% of Americans participating in any type of camping respectively. Literature suggests that this is due to the notoriety and brand loyalty associated with national parks (Mayo, 1975; Reinius & Fredman, 2007; Stephens, Leonard, & Smolder, 1989; Weiler, 2006). Weiler and Seidl (2004) find this phenomenon has an even greater impact on international visitors than residents.

From 2013 to 2017, visitation to NPS sites in the United States increased by 20.9% (National Park Service, 2018a). An increase in demand for campsites has mirrored this significant uptick in visitation. Over the same time span, the amount of tent campers in NPS sites has increased 24.7%, and RV camping has increased 24.5% (National Park Service, 2018b). Yosemite National Park provides a vignette of how this excess demand presents itself. In the summer months, campers arrive before dawn to get in line at Camp 4 with hopes of securing a campsite. It is not uncommon for visitors to arrive in the afternoon prior to their anticipated stay in order to stake their spot in line (Almond, 2017). Other “first come, first served” campgrounds in the park fill by mid-morning (Yosemite National Park, 2018). Media outlets report this trend across national parks (Siegler, 2016).
To aid visitor planning and reduce behavior like that reported in Yosemite, many parks have instituted online reservation systems through recreation.gov. This website allows campers to reserve campsites as they would hotel rooms. Since its inception, the number of parks using the service has grown from 12 in 2006 to 57 in 2018. Walls, Wichman, and Ankney (2018) found that most campgrounds listed on recreation.gov were filled to capacity during the height of the summer season and parks near large population centers filled on weekends during shoulder seasons. These indicate a great need in obtaining accurate forecasting for better management and resource allocation for campgrounds in national parks.

**Forecasting Methods for Tourism and Hospitality**

The importance of accurate demand forecasting has been emphasized repeatedly in the tourism literature for policy development and implementation (Goh & Law, 2011; Hassani, Silva, Antonakakis, Filis, & Gupta, 2017; Li, Song, & Witt, 2005). A list of review articles have covered the forecasting methods applied for tourism demand (Goh & Law, 2011; Jiao & Chen, 2018; Li et al., 2005; Peng, Song, & Crouch, 2014; Song & Li, 2008; Witt & Witt, 1995; Wu et al., 2017). Six review articles, from Witt and Witt (1995) to Jiao and Chen (2018), provide an extensive overview of the literature on the topic of tourism demand modeling and forecasting for the past five decades.

Prior to the early 2000s, most of the studies employed econometric models (Witt & Witt, 1995) and structural time-series model with explanatory variables, such as autoregressive (integrated) moving average model with explanatory variables (Li et al, 2005). Song and Li (2008) and Goh and Law (2011) reviewed 119 and 155 articles respectively up to the first decade of 2000s. Wu et al. (2017) and Jiao and Chen (2018) are the latest review articles that each explored 171 articles between 2007 and 2015 and 72 between 2008 and 2017, respectively. All four reviews report that a wide variety of time-series models—non-causal time series models, causal econometric approaches and artificial intelligence (AI)-
based methods—has been applied in tourism demand forecasting studies.

There is no evidence for a universally superior forecasting method when individually examined (Goh & Law, 2011; Song & Li, 2008; Wu et al., 2017). However, according to Peng et al. (2014)—a meta-analysis of tourism forecasting methods accuracy—a dynamic econometrics model was the most popular and had the highest accuracy across the 2,584 studies examined, followed by AI-based and advanced time-series models.

Jiao and Chen (2018) also identify AI as a rapidly developing area of research and combination models as one of the emerging trends in tourism forecasting. Their study undertakes both the AI-based measures and the hybrid model that combines AI and advanced time-series models. For AI-based methods, artificial neural network, support vector regression, fuzzy time series, grey theory, genetic algorithms, and combinations have previously been used (Jiao & Chen, 2018; Wu et al., 2017).

However, there is lack of research that applied k-nearest neighbors (KNN) algorithm even though it is one of the simplest and effective AI approaches (Noersasongko et al., 2016). KNN is a nonparametric machine learning method that predicts a value based on its k most similar cases (Díaz & Mateu-Sbert, 2011; Martínez, Frías, Pérez, & Rivera, 2017). Diaz and Mateu-Sbert (2011) and Noersasongko et al. (2016) are the only articles identified by Jiao and Chen that have adopted KNN to forecast tourism demand. The former forecasted daily air arrivals in Mallorca Island, Spain; the latter compares Backpropagation Neural Network, KNN, and Multiple Linear Regression to find the optimal predictor for international tourist arrivals in the district of Central Java, Indonesia.

Both studies used recursive strategy where first a single value is forecasted based on the training data and the following prediction’s training data includes the first forecast. Hence, the recursive approach can suffer from accumulated errors (Martínez et al., 2017). Therefore, in this study, we take a multi-input multi-output (MIMO) strategy where multiple
forecasts are predicted simultaneously. MIMO can be more efficient compared to recursive strategy as it allows greater number of forecasts with less training data, and avoids cumulating errors (Martínez et al., 2017). KNN with MIMO is expected to perform well with predicting the RIDB data as similar reservation patterns repeat each year for many campsites. Thus KNN is selected to complement the existing methods.

Advanced time-series models are used to assess and complement the accuracy of the machine learning methods. We use non-causal forecasting methods, which usually employ time series analysis to capture a particular pattern or behavior. This includes different types of exponential smoothing (ETS) models such as Holt-Winters seasonal method and autoregressive moving average (ARIMA) models such as seasonal ARIMA (SARIMA) (Song & Li, 2008; Wu et al., 2017). We conduct SARIMA and ETS with Holt-Winters seasonal method as advanced time-series methods. We also run simple moving average (MA) as the base model. We also combine machine learning and advanced time-series forecasting tools to achieve higher accuracy. Specifically, we develop a model that is an ensemble of ETS with Holt-Winters seasonal method, SARIMA, NNAR, and KNN.

**Uniqueness of Campground Forecasting**

Previously, forecasting national park campground demand has been carried out through relatively costly measures such as on-site and telephone surveys (Snepenger & Karahan, 1991) or through basic linear trend analysis (Marin-Pantelescu, 2015). The survey is limited in that the collection of the data has to be repeated for each prediction, costing money and time, and that the data can only be used to forecast demand for a specific location during a specific period. Moreover, such surveys only intercept a small proportion of the visitors and do not take into account the information from previous years. Span (2017), a master’s thesis, focused on dynamic pricing for revenue maximization, but did not develop measures to forecast future demand. Therefore, there is a need for the development
of more robust forecasting measures.

The lack of forecasting concerning campgrounds is underscored by the inherent difference between campgrounds and hotels. First, different from the tourism and hospitality industry, American national parks were established under what is known as the dual-mandate. The NPS is to manage their parks in a way as to provide for the enjoyment of the people while leaving the parks unimpaired for future generations (Sax, 1980). For this reason, merely adding more campsites to meet increased demand is not easily accomplished. Deng, King, and Bauer (2002) note that as nature-based tourism continues to grow, pressure on the resources of national parks will increase. The impact of building new roads and constructing more tent pads must be carefully evaluated. Such processes can take decades to accomplish. This is in sharp contrast with the hotel industry, where new hotels can be built and existing hotels can expand their room supply if needed. Even though sometimes one campground can stretch its inventory by temporary adding a few unclaimed and undesignated sites, this is more likely to be rare cases. Thus, many parks are at full capacity during peak seasons and visitors are left with unmet demand for campsites. The seasonality of campgrounds underscores their uniqueness when contrasted with forecasting hotel demand. Many national park campgrounds close completely for the winter months, leaving long breaks in reservation data.

In addition, campers and hotel guests are distinct subsets of tourists. Though very little research has examined the differences between them, the available literature suggests the tourists to hotels versus campgrounds vary in their demographics, length of stay, and demand for visitation (Becken & Gnoth, 2004; Eagles & Wade, 2006; Poudyal, Paudel, & Tarrant, 2013). For example, campers are most likely to be recreation tourists, unlike hotel guests which also include a lot of business travelers and family and friends-visiting tourists. Campgrounds and hotels are often also classified as different categories of lodging types and
possess different characteristics (Foris, 2014). For instance, unlike hotels that usually open
year-round, most campgrounds have an off-season where they are closed and inaccessible
(Manning & Powers, 1984; Schreuder, Tyre, & James, 1975). The existence of an off-season
can skew visitation patterns (Manning & Powers, 1984).

Therefore, the forecasting of campground demand should be conducted separately
from general tourism demand forecasting methods to locate the most suitable technique that
takes into account the different characteristics of campgrounds and their visitors. Accurately
forecasted demand allows efficient use of the campgrounds, which are perishable goods—
unoccupied campsites cannot be stored and resold (Frechtling, 2012). Moreover, efficient
forecasting aids the policymakers, in our case the NPS administrators, to develop adequate
policies to maintain economic and environmental sustainability, and to make decisions for
infrastructure investment (Wu et al., 2017). A proven and generalizable forecasting method
that utilizes past campsite demand data will not only contribute to the existing literature but
also benefit the policymakers and the industry. The following section introduces the data and
the methods used to forecast the campsite demand.

Data and Methods

Campground reservation data for this forecasting analysis covered the years from
2007 to 2017 and was acquired from the Recreation Information Database (RIDB). The
RIDB provides publicly available historical data which record individual reservation data for
destinations including federal recreation areas, campsites, tours, and more. The data contain
prior or on-site reservations made for campgrounds managed by the federal government
(Supak, Brothers, Ghahramani, & Van Berkel, 2017).

Table 1 shows a summary of the campground reservation data. However, the data are
not an exhaustive list of all campgrounds in national parks, and the reservation availability
may vary across years and seasons. Eleven out of 31 parks have 100% of their campsites on
recreation.gov, and 28 have more than 80% as of November 2018 by comparing to statistics on NPS.gov website. As confirmed by a previous study, examining the reservation data from the RIDB can provide useful insights for understanding the overall campground reservation patterns (Walls, Wichman, and Ankney, 2018).

As the purpose of the current study is to forecast the demand for the campsites in national parks, only the reservations that are made for campsites operated by the NPS were examined. Those with erroneous information are removed (Supak et al. (2017): the current study observed and eliminated reservation records of which the start date was later than the end date or did not have an end date. This process reduced the dataset to 3,685,260 reservation records within 108 campgrounds associated with 32 national parks. In total, these campgrounds had 8,363 campsites run by the NPS of which 7,575 were registered on recreation.gov (Table 1).

A visual examination of the campground metadata indicates that each campground has a unique occupancy, demand, and set of administrative rules that govern what portion of the inventory are posted on recreation.gov, how far in advance they can be booked, and what vehicle types are allowed. Therefore, reservations were aggregated by individual campground and date to create a longitudinal dataset. Data points for every date between the first start date and the last end date were filled. If there was no reservation record for any dates, a value of zero was given.

The final dataset contained 346,539 data points from 108 campgrounds. For each campground, the dataset records the average values for length of stay, booking horizon (the length of days between reservation and check-in), fee, revenue, number of people, and occupancy (percentage of campsites occupied) for each day between the start date of the first reservation to the end date of the last reservation. The summary statistics are shown in Table
2. To avoid any misleading conclusions, outliers were excluded when calculating summary statistics (see footnotes in Table 2).

Park and Campground Level Data

The campground reservation data were compiled across 32 national parks by aggregating all reservation data of one or more campgrounds in the same national park. The parks range geographically from the Everglades of South Florida to the Katmai Valley of the Alaska Peninsula (Figure 1).

As mentioned above, each park and campground manifest very different characteristics (See Figure 2). Thus, we chose five campgrounds from different national parks for more detailed analysis and forecasting: Big Meadows Campground in Shenandoah National Park, Elkmont Campground in Great Smoky Mountains National Park, Mather Campground in Grand Canyon National Park, Moraine Park Campground in Rocky Mountain National Park, and Upper Pines Campground in Yosemite National Park. We selected the campgrounds based on their relatively high percentage of campsites listed on recreation.gov. Though Blackwoods Campground in Acadia National Park and Flamingo Campground in Everglades National Park were initially included in the top five based on the number of campsites, they were excluded due to a large percentage of erroneous data.

All five campgrounds boasted a. Just Mather Campground has around 95% of campsites listed (94%). As reported in Table 3, average occupancy ranged from 49.4% (Big Meadows) to 89.4% (Moraine) among campgrounds. Average booking horizon ranged from 60.4 days (Big Meadows) to 84.2 days (Upper Pines). Average length of stay ranged from 2.15 days (Mather) to 4.02 (Elkmont).

The five campgrounds selected, and most large NPS campgrounds, are designed in the
Meinecke style in that they are composed of numerous loops fringed with campsites and “parking spurs” that commonly feature a picnic table, tent pad, and fire ring (Hogue, 2016). A campsite might also have access to electricity and water hook-ups for RVs. Each loop usually has at least one bathroom and water spigot. This common design allows us to treat these geographically-varying campgrounds in the same way.

For four of the five campgrounds, we were able to observe average occupancy greater than 100%. This is due to a unique quality of campgrounds, as opposed to hotels. Campsites can be split into smaller sites when a campground is overbooked, or temporarily adding undesignated sites if a special reservation is required (A. Blietz, personal communication, October 1, 2018).

Methodology

We conducted six forecasting methods as it is one of the first attempts to rigorously forecast campsite demand and previous literature on forecasting in the tourism domain has shown that there is no single method with definite superior performance (Hassani et al., 2017). The current study selected four forecasting models that have been widely applied and found to be effective in the tourism literature (Frechtling, 2012; Hassani et al., 2017; Song & Li, 2008) as well as two relatively new approaches to explore and compare the predictive ability and identify the most accurate model: Moving average (MA), ETS with Holt-Winters seasonal method, SARIMA, NNAR, KNN, and the combination of ETS, SARIMA, NNAR and KNN. Unit root tests were carried out prior to the analyses to examine seasonality.

Neural network autoregression (NNAR)

Neural network autoregression (NNAR) is one of the AI-based forecasting method that conducts forecasting using lagged time series as the weight (Hyndman & Athanasopoulos, 2018). It is the most frequently applied AI-based method in the tourism and
hospitality domains (Wu et al., 2017) and has previously provided reliable forecast performance, especially for discontinuous data (Kon & Turner, 2005).

For this study, we employ multilayer feed-forward network with seasonality. The multilayer feed-forward network assumes that the final output is a non-linear weighted combination of the original inputs (Hyndman & Athanasopoulos, 2018). The NNAR model creates hidden inputs that are the linear weighted combination of the original inputs, then the hidden inputs are again linearly combined to formulate the final output. The weight for each input is learned from data. That is, the weight first starts with a random number but continues to adjust based on the training data.

With seasonality, NNAR model can be presented as $\text{NNAR}(p, P, k)_m$ where $p$ is the number of lagged inputs, $P$ the number of the last observed values from the same season, and $k$ the number of hidden inputs. For instance, the inputs for $\text{NNAR}(1,1,2)_{12}$ is $(x_{t-1}, x_{t-12})$ with two hidden nodes. $P$ and $m$ was set to 8 and 12 respectively whereas $p$ was based on the fitted value of optimal linear model using the seasonally adjusted data. We set the number of iteration to 100.

**K-nearest neighbors (KNN)**

KNN is a nonparametric method that is often used for classification and regression. It identifies the most similar data points based on the explanatory variables for classification or prediction. For univariate time-series data as in this study, KNN utilizes the lags of the forecast variable as the explanatory variables (Martínez et al., 2017). Our current study employs KNN with multi-input multi-output (MIMO) strategy to forecast campsite demand.

According to Taieb, Bontempi, Atiya and Sorjamaa (2012), MIMO strategy forecasts a vector of $h$ data points $[y_{t+h}, \ldots, y_{t+1}]$ using $d$ lags $[y_t, \ldots, y_{t-d+1}]$. MIMO allows consideration of the stochastic dependencies between future values, which is a limitation of single-output mapping. For this study, the number of lags is set to 12 as the campsite
reservation manifests seasonality of 12, where the pattern repeats every one year. The $k$ is set to three, which is the square root of the number of training instances, 12. The Euclidean distance $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ is used to measure the distance between the data points.

**Moving average (MA)**

The forecasting efficiency of the Moving average (MA) model is used as the baseline model. MA method is related to past errors. MA(q) used in the current study is a moving average model that assumes nonzero autocorrelations for the first $q$ lags. That is,

$$x_t = \mu + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

where $\mu = E(x_t)$ and $e_t \sim iid \mathcal{N}(0, \sigma_e^2)$, meaning that the $e_t$ are independently and identically distributed, each with a normal distribution having a mean of zero and variance $\sigma_e^2$. Simple MA(q) model is applied to provide the baseline to compare the accuracy of all other models. Autocorrelation function (ACF) and campground operating policies were taken into consideration to determine the lag $q$. The ACF manifests non-zero autocorrelations for significant lags. The ACF between $x_t$ and $x_{t-q}$ is

$$\mu(q) = \frac{Covariance(x_t, x_{t-q})}{Variance(x_t)}.$$  

For this study, moving average (MA) with lag 3 was selected: $x_t = \mu + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3}$. That is, we predicted the occupancy at time $t$, $x_t$, using the mean occupancy, $\mu$, and the past three error terms, $\theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3}$. The results from MA(3) served as the baseline for accuracy comparison.

**Seasonal ARIMA (SARIMA)**

Auto-regressive integrated moving average (ARIMA) model is one of the most popular time-series and forecasting model used in hospitality and tourism literature (Hassani et al., 2017; Song & Li, 2008). However, ARIMA fails to incorporate the seasonality, which is one of the dominant characteristics of tourism industry (Song & Li, 2008). Campgrounds manifest even stronger seasonality due to their subjectivity to weather and operational
constraints. Many campsites close during winter season, mostly between November and March. Our dataset had 108 campsites, of which the majority (69%) were closed during winter and early spring—as there will be no occupancy during closed period. The Seasonal ARIMA (SRIMA) model investigates both non-seasonal and seasonal factors using a multiplicative model and has gained popularity over the recent years (Song & Li, 2008). Simply put, the model is as follows:

\[ ARIMA(p, d, q) \ast (P, D, Q)_S \]

where \( p \) = non-seasonal AR order, \( d \) = non-seasonal differencing, \( q \) = non-seasonal MA order, \( P \) = seasonal AR order, \( D \) = seasonal differencing, \( Q \) = seasonal MA order, and \( S \) = time span of repeating seasonal pattern. The campground demand exhibits time span of \( S = 12 \) for seasonality and the previous year’s data would help us predict this year’s month’s value.

To determine an appropriate model for each campground, we employed optimized Box-Jenkins ARIMA model obtained from auto.arima() function of forecast package in the R software (Hyndman & Athanasopoulos, 2018):

The auto.arima() function in R uses a variation of the Hyndman-Khandakar algorithm (Hyndman & Khandakar, 2008), which combines unit root tests, minimization of the AICc and MLE to obtain an ARIMA model. (Hyndman & Athanasopoulos, 2018, ch 8.7)

The order of the AR model, the degree of differencing, and the order of the MA model were determined by repeated Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Osborn-Chui-Smith-Birchenhall (OCSB), Phillips-Perron (PP), Hylleberg-Engle-Granger-Yoo (HEGY), and Canova-Hansen (CH) tests; and minimizing the corrected Akaike information criterion (AICc) (Hurvich & Tsai, 1989).
Exponential smoothing (ETS)

Though seasonality is inevitable due to the operational style of the campgrounds, often smoothing out the seasonality can help better detect the trend. Exponential smoothing (ETS), developed in the late 1950s, uses the average of the past observations with weights that exponentially decrease (Hyndman & Athanasopoulos, 2018). Of various ETS models, we employ Holt-Winters' seasonal method. Based on Holt (2004), which incorporated trend into simple exponential smoothing (SES), Holt (2004) and Winters (1960) developed ETS method that included level, trend, seasonality, and smoothing (Hyndman & Athanasopoulos, 2018). Holt-Winters' method has two variations: additive and multiplicative. The additive model assumes a constant degree of seasonality; the multiplicative method assumes seasonality with a multiplicative relationship between the trend, seasonality, and irregularity in the series.

The current study’s analysis assumed additive errors, no trend, and additive seasonality for all campgrounds: ETS(A,N,A). The appropriate model was selected based on AICc. The function utilizes normalized version of ETS(A,N,A):

\[
\begin{align*}
    x_t &= l_{t-1} + s_{t-12} + \epsilon_t \\
    l_t &= l_{t-1} + \gamma \frac{\epsilon_t}{12} \\
    s_t &= s_{t-12} + \gamma \epsilon_t - \gamma \frac{1}{12} (\epsilon_t + \epsilon_{t-1} + \ldots + \epsilon_{t-12})
\end{align*}
\]

(6)

where \( l_t \) is the level (or the smoothed value) of the series at time \( t \) and \( s_t \) is the seasonal component.

Forecast accuracy assessment

To assess the accuracy of the forecasted value \( f_t \) against the original value \( x_t \), we examine scale-dependent errors such as mean error (ME = \( \frac{\sum_{t=1}^{n} (x_t - f_t)}{n} \) = \( \frac{\sum_{t=1}^{n} \epsilon_t}{n} \)), mean absolute error (MAE = \( \frac{\sum_{t=1}^{n} |\epsilon_t|}{n} \)) and root mean squared error (RMSE = \( \sqrt{\frac{\sum_{t=1}^{n} \epsilon_t^2}{n}} \)) and percentage errors such as mean absolute percentage error (MAPE = \( \frac{1}{n} \sum_{t=1}^{n} \frac{|x_t - f_t|}{x_t} \)). ME, MAE,
MAPE, and RMSE are popular measures to assess accuracy of the forecasting models (Wu et al., 2017). Since our dataset contain zero occupancy values for closed seasons, some of the MAPE values were infinite due to the denominator being zero. Therefore, we manually calculated MAPE after assigning zero to closed months.

**Results**

The current study focused on forecasting campground occupancy (i.e. the percentage of campsites occupied) as we concluded that occupancy is the most pressing concern of the campgrounds based on our conversation with national park officials. To calculate occupancy, we divided the number of campsites (Product ID in the RIDB database) reserved by the total number of campsites, which we manually collected from the recreation.gov website. Though we have daily reservation data, forecasting at the monthly-level was considered more appropriate and useful. Figure 3 graphically presents the monthly occupancy for all five campgrounds. We observe clear seasonality and unique pattern for each campground.

![Figure 3 about here](image)

**Unit Root Tests of Stationarity**

Four types of unit root tests were conducted (Table 4): Augmented Dickey-Fuller (ADF) based on Banerjee et al. (1993, p. 170-171, Table 6.3) and McKinnons (1996), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Osborn-Chui-Smith-Birchenhall (OCSB), and Canova-Hansen (CH) tests. The former two are unit root tests and the latter two are seasonal unit root tests. The null hypothesis of ADF is that a series is non-stationary (Said & Dickey, 1984). KPSS, on the other hand, is a stationarity test (Kwiatkowski, Phillips, Schmidt & Shin, 1992) but for this study we set the null hypothesis to stationary for less confusion. OCSB tests the null hypothesis that a seasonal unit root exists (Osborn, Chui, Smith & Birchenhall, 1988) whereas the CH tests the non-existence of seasonal unit root (Canova & Hansen, 1995). The test results in Table 4 indicate that for the
first three campgrounds, there is no unit root but seasonal differencing is required. Moraine and Upper Pines show conflicting testing results. In order to keep the analysis consistent, we apply seasonal differencing to all campgrounds.

For SARIMA, we conducted an additional unit root test of Phillips-Perron (PP) (Phillips & Perron, 1988), and two additional seasonal unit root tests of seasonal strength (Wang, Smith, & Hyndman, 2006) and Hylleberg-Engle-Granger-Yoo (HEGY) (Hylleberg, Engle, Granger, & Yoo, 1990) to determine differencing.

| Table 4 about here |

**Forecasting results**

Six forecasting methods were used in the current study: MA, SARIMA, ETS, NNAR, KNN, and a combination model that averages the forecasts from the last three methods. For each method, we conducted three, six, and twelve months forecasting. Respectively, we predicted occupancy rates for June–August 2017, March–August 2017, and September 2016–August 2017 based on the previous data and compared them to the actual values. All analyses were performed using R Statistical Software (version 3.5.1).

We did not restrict SARIMA, ETS, and NNAR to a single model for all campgrounds considering the different characteristics of the campgrounds, such as operational rules and geographical locations. We used AICc and ADF, CH, HECY, KPSS, PP, and OSCB tests to determine SARIMA’s non-seasonal and seasonal differencing and AR and MA orders. For Big Meadows and Moraine Park campgrounds SARIMA(0,0,0)(0,1,0)\text{12} model was selected. For Elkmont and Mather Campground, SARIMA(0,0,1)(0,1,0)\text{12} model was selected; and SARIMA(1,0,1)(0,1,0)\text{12} model for Upper Pines Campground. All campground occupancy data were seasonally differenced and those of Elkmont and Mather took one forecast error into account; Upper Pines included one lagged occupancy in addition to the error.

ETS(A,N,A) model was selected based on AICc for all campgrounds, which assumes
additive errors, no trend, and additive seasonality. Non-seasonal lags for NNAR were
determined using AIC, and the number of hidden inputs was set to half of the number of
inputs plus one. The number of seasonal lags was given the value of eight to take into account
the maximum number of lags.

As mentioned in the methods, we used mean error (ME), mean absolute error (MAE),
mean absolute percentage error (MAPE), and root mean squared error (RMSE) to assess
accuracy (Wu et al., 2017). For some of the closed seasons, MAPE was manually calculated
taking into account zero occupnacy.

Tables 5 and 6 shows the accuracy for each forecasting model. Table 5 presents the
forecasting accuracy per campground for six months ahead. We boldfaced the method with
lowest average error value, the last column of Table 5. There is no universal measure that
performs the best for all the campgrounds. We believe this inconsistency is due to the
different characteristics of the campgrounds. Except for MA, each model performs the best
for at least one campground, which is why the combination model comprised of SARIMA,
ETS, NNAR, and KNN.

Table 6 is a table of the averaged MAPE for each method for three, six, and twelve
months forecasts. SARIMA has the lowest error rate for three months ahead forecast whereas
the combination model performs the best for six and twelve months ahead. Figure 4
demonstrates the graphic comparison between the actual reservation data and the forecasts of
six and twelve months for SARIMA and the combination. The black line represents the
original reservation data, the blue line represents the SARIMA forecasted value, and the
orange line that of the combination method. For all five campgrounds, we are able to observe
that the performance from all forecast methods is satisfactory.
Discussion

The results of the current study demonstrate that SARIMA, ETS, NNAR, and KNN can provide reasonable forecasting accuracy as measured by popular error rate indices. MA is the only forecasting method where the average percentage errors consistently exceeded 10%, thus, it is not suitable for forecasting purpose. The ensemble method, that is, the average of four most popular methods, provides the best accuracy. SARIMA is the second model among the alternatives, which might provide a cost-effective method for forecasting campground demand. In retrospect, the strong seasonality may render the MA method unfeasible, since it does not consider data points from the past year. ETS also considers seasonality in the data by Holt-Winter’s approach; NNAR includes seasonality by considering past 12 months; KNN looks for similar values in the past data points which are likely to be the same month in the past time period.

Even though ensemble method provides the most accurate model but it is time-consuming to construct by modeling with three different methods; thus, SARIMA is the best cost-effective approach and suitable for practical adoption. Individual campgrounds could adopt SARIMA in forecasting their monthly occupancy.

It should be noted that we anticipated the Great Recession of 2008 would make for somewhat less predictable campground reservation trends, given our window of analysis. However, the trends from these campgrounds reveal that demand was seemingly unaffected by the economic event. This follows the Poudyal et al. (2013)’s finding that front country camping in national parks had more inelastic demand during the recession. McIntosh and Wilmot (2011) also argued that campsites in national parks are inferior goods, as opposed to hotel rooms, despite their appeal. The likelihood of staying at a campground increases as disposable income decreases. This is not in line, however, with the research of Walls, Wichman, and Ankney (2018). A little more details here on how?
Implications

Due to the increasing demand for campsites, forecasting campers’ demand has direct applications for campground managers, administrative policymakers, other tourism sectors, and campers themselves. Knowing the expected occupancy and the expected fill date of a particular campground have outstanding relevance to trip planning. Given forecasts, park managers can communicate suggested planning horizons via news releases on their website, or through social media outlets. Further, managers can adjust the allowable booking window to meet the predicted planning horizon of their campers. If the suggested planning horizon is communicated to the public through a consistent, accessible means, consumers will be able to gain an understanding of the amount of pre-planning required for a particular campground and make informed destination choices. Gaining a better understanding of future campsite demand and sharing that information throughout the tourism supply chain can reduce uncertainty and demand shocks up the supply chain from campers to park managers (Yu, Yan, & Cheng, 2001). In this way, park concessionaires can adjust their food preparation, park rangers can adjust their staffing, and guide services can vary their offerings based on predicted demand.

Of course, the finding that campground demand is highly variable according to seasonality creates additional problems of labor allocation and resource use. Smoothing campsite demand, therefore, must also be considered as a potential application of forecasting. In the realm of federal land management agencies, such as the NPS, altering demand variability is likely the task of policymakers. Other governments have employed market solutions to smooth demand. Parks Canada has implemented a dynamic pricing scheme, adding a premium to their campsite prices during peak seasons (Beaman, Hegmann, & Duwors, 1991).

Similarly, Vermont State Parks have experimented with differential campsite pricing
based on occupancy rates (Bamford, Manning, Forcier, & Koenemann, 1988). Inherent to this discussion of fee variation on public lands is the question of social equity, however. Park, Ellis, Kim, and Prideaux (2010) examined the issue of raising campground fees within the context of the impact on low-income campers. They found that visitors were more accepting of the social implications of higher user fees when they are established after extensive public input, predicting both price acceptability and social equity. Parks might also consider presenting certain amenities in campgrounds as add-on services, available with an extra fee. Ostergren, Solop, and Hagen (2005) found this strategy to be preferred across income groups.

Alternatively, forecasts can be used in non-market approaches to the allocation of campsites such as lottery systems. For high demand activities such as rafting through the Grand Canyon, viewing the synchronous firefly display in the Great Smoky Mountains, and backcountry camping in Glacier Park, the NPS has implemented a lottery system (often through recreation.gov) to distribute limited permits to scarce recreational opportunities in a more equitable fashion (Manning, Anderson, & Pettengill, 2017). Park managers seem to favor this mode of allocation over price shifts because it more directly conforms to preset ecological and social carrying capacities of protected areas (e.g. Pettebone et al., 2013).

Limitations

The RIDB dataset is somewhat limited in that it only includes data for campsite reservations made through recreation.gov or entered by campground managers for same-day reservations. Though it is the perception of the data managers that same-day reservations are rarely being inputted (S. Gregory & E. Levine, personal communication, September 17, 2018). Agency officials estimate that roughly thirty to fifty percent of all NPS managed campgrounds are reserved through recreation.gov (S. Gregory & E. Levine, personal communication, September 17, 2018). This excludes those campgrounds operated by licensed concessionaires. Additionally, each campground operates under unique “business rules”
which include the allowed booking window, maximum length of stay, resale policy, and vehicle restrictions (S. Gregory & Levine, personal communication, September 17, 2018). Each campground also has unique rules concerning the management of overbooked occupancy and saving “hidden” campsites for last minute arrivals (A. Blietz, personal communication, October 1, 2018). The inconstancy of these rules combines to create a lack of consistency in the data.

**Conclusion**

Camping in iconic national parks, such as those found in the United States, is a rapidly growing segment of tourism. At the same time, supply of campsites remains largely fixed. For the managers of these campgrounds, it is necessary to anticipate their future demand in order to provide proper and timely information, efficient booking windows, and staffing as a means of maximizing the tourist experience. Our research suggests that the combined method proves to be the best in forecasting the unique, highly seasonal, and site-specific demand of campsites in national park campgrounds. Our results underscore the impact of seasonality on campsite demand and its impact on forecasting.

**Future Research**

RIDB datasets offer many more metrics to measure the relative demand for campsites. Future research should assess the validity of booking window, occupancy, and other measures as significance indicators, thus filling a longstanding need for a measure of demand-driven significance in the park and tourism literature.

This study has revealed the varying trends and limitations experienced when examining RIDB data at different scales. We advocate that future research transcend the campground scale and examine the relative demand for individual campsites within a campground. Through this lens, the effect of ecological and managerial amenities have on demand could be measured. As Hogue (2016) argues, the online campsite reservation
capability has dramatically changed the camping tourist’s experience, marking a new era in camping’s history even, where the real and virtual aspects of the experience merge into a year-round activity and the focus is on the campsite rather than the campground. With booking windows reaching six or more months in advance, the activity of selecting the perfect site becomes a leisure activity unto itself.

Finally, our successful application of forecasting to campgrounds opens up the possibility of applying these methods to other increasingly popular components of outdoor recreation-based tourism. Forecasting demand for permits for popular tours, rafting trips, and hikes could be attempted as a means of helping park administrators better allocate these exclusive opportunities.
References


