Using Social Media Data to Understand Park Visitation

Diane Kuehn, James Gibbs, Harrison Goldspiel, Brannon Barr, and Lilly Stradtman
SUNY College of Environmental Science and Forestry, Syracuse, NY

Alden Sampson and Marshall Moutenot
Upstream PBC

Introduction

Forest-based recreation and tourism are important and growing components of the Northern Forest Region's economy and quality of life (Figure 1). Tourism promotion agencies, state and national forest management entities, and academics rely on visitor use data to guide recreation and tourism planning efforts; however, existing empirical visitor use data are often not detailed, not available, expensive to collect, and/or limited in geographic scope.

Goal: To explore the utility of data "mined" from social media for quantifying forest-based tourism in the Northern Forest Region (NFR).

Objective: To determine if the number of images posted by the public can serve as a valid proxy for estimating visitor use.
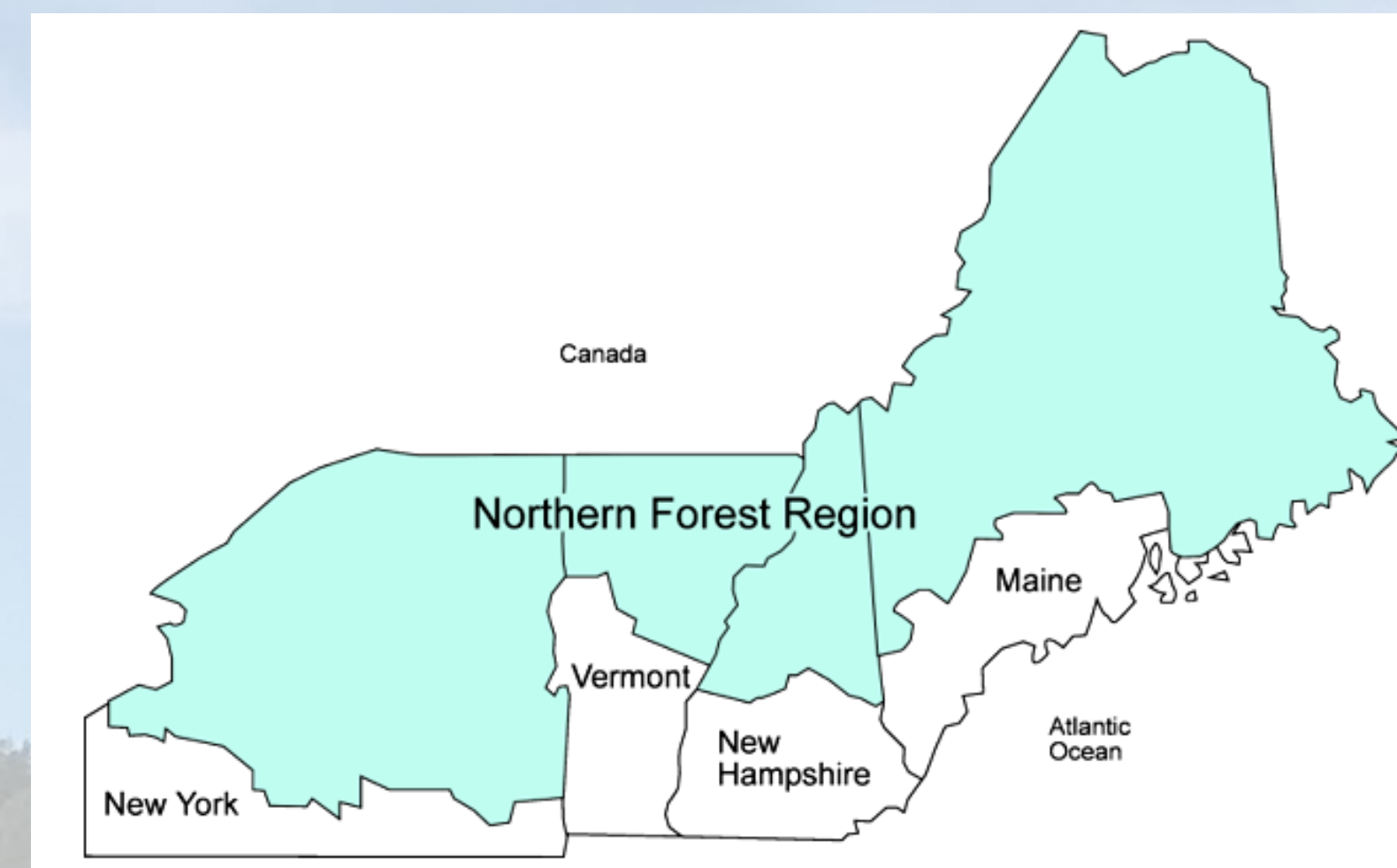

Figure 1. Northern Forest Region of NY, VT, NH, & ME (adapted from a map by the Conservation Advisory Service, 1994).

Methods

Summer park admissions data for Maine, New Hampshire, Vermont, and New York were obtained from the park agency in each state for the years 2012 through 2016. The image identification program Clarifi was used to identify the number of images posted to Flickr in the NFR for each park; the time and location of each photograph was identified, and the mean number of images per day for June through September of each year for each park were calculated. Data were also compiled for the number of amenities (0 to 9), size (km^2), road access (1) or boat access only (0), and destination level (primary = 3, secondary = 2, scenic/ service hub = 1) of each park. A stepwise linear regression was used to identify the relationships between summer attendance (dependent variable) and the independent variables for all states combined and for each state separately.

Results

Although the regression analysis for the Northern Forest Region as a whole was not significant, the regressions for each state (i.e., ME, NH, VT, and NY) were significant. Different social media data were tested as independent variables (e.g., mean images per day; mean photographers per day); the variable significant for most of the states (excluding VT) was mean images per summer day. Results of the state regressions are shown in Table 1.

Table 1. Regression analysis results in the NFR.

State	Dependent Variable	Independent Variables	Beta	p-value	Partial correlation coefficient	VIF
Maine F = 669.506 p < .001 R^2 = 0.970 n = 45	Summer attendance	Area	0.734	< .001	0.926	2.958
		Mean images/summer	0.291	< .001	0.696	2.958
New Hampshire F = 152.402 p < .001 R^2 = 0.909 n = 50	Summer attendance	Mean images/summer	0.960	< .001	0.931	1.545
		Number of amenities	-0.184	0.002	-0.429	1.644
		Destination level	0.158	0.007	0.386	1.547
New York F = 237.966 p < .001 R^2 = 0.746 n = 165	Summer attendance	Area	0.164	< .001	0.306	1.026
		Mean images/summer	0.822	< .001	0.850	1.026
Vermont F = 53.395 p < .001 R^2 = 0.648 n = 91	Summer attendance	Destination level	0.379	< .001	0.527	1.057
		Number of amenities	0.673	< .001	0.732	1.117
		Road access	0.253	< .001	0.370	1.144

Conclusion

The regressions revealed that the independent variables were not universally significant among all states in the NFR, likely due to park management differences in each state. Additionally, the time, expertise, and resources needed to collect and analyze social media data may preclude their use by park managers for attendance estimation purposes. In summary, though significant relationships were found among summer attendance and the independent variables included for each state in the NFR, continued use of existing park attendance estimation methods will likely be needed in the near future, until estimation techniques utilizing social media data become less time-consuming and expensive to implement.