

Das Potential von Daten aus sozialen Medien für die Erforschung der Walderholung



Pilotprojekt als ergänzendes Testmodul zum Projekt
«WaMos meets LFI» (WML)

Flurina Wartmann
Manuel Bär
Ross Purves
Marcel Hunziker

Im Auftrag des Bundesamtes für Umwelt (BAFU)
Interner Bericht an den Auftraggeber

Impressum

Auftraggeber:

Bundesamt für Umwelt (BAFU), Abteilung Wald, CH-3003 Bern
Das BAFU ist ein Amt des Eidg. Departements für Umwelt, Verkehr, Energie und Kommunikation (UVEK)

AutorInnen:

Dr. Flurina Wartmann, Eidg. Forschungsanstalt WSL, flurina.wartmann@wsl.ch
Manuel Bär, MSc., Geographisches Institut UZH, manuel.baer@uzh.ch
Prof. Dr. Ross Purves, Geographisches Institut UZH, ross.purves@geo.uzh.ch
Dr. Marcel Hunziker, Eidg. Forschungsanstalt WSL, marcel.hunziker@wsl.ch

Begleitung BAFU

Clémence Dirac Ramohavelo, Sektion Waldleistungen und Waldpflege

Hinweis

Dieser Bericht wurde im Auftrag des BAFU verfasst. Für den Inhalt sind allein die AutorInnen verantwortlich.

Zitiervorschlag

Wartmann, F.; Bär, M.; Purves, R.; Hunziker, M. (2018): Das Potential von Daten aus sozialen Medien für die Erforschung der Walderholung. Pilotprojekt als ergänzendes Testmodul zum Projekt «WaMos meets LFI» (WML). Im Auftrag des Bundesamtes für Umwelt (BAFU). Interner Bericht an den Auftraggeber. Birmensdorf, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft, WSL.

Titelbild

Jan Remund auf Flickr (CC BY 2.0) <https://tinyurl.com/y9935rc2>

Inhaltsverzeichnis

Impressum	1
Zusammenfassung.....	3
Einleitung	4
Von Tweets und Tags: Daten aus sozialen Medien für die Forschung	5
Literaturreview zu sozialen Medien für die Erforschung von Erholungsnutzung	7
Besucherhäufigkeit	7
Besucherverhalten.....	9
Präferenzen und Wahrnehmung	11
Testweise Analyse von Daten aus sozialen Medien zur Erforschung von Erholungswäldern in der Schweiz	13
Verfügbarkeit von Daten aus sozialen Medien in Schweizer Wäldern	13
Stichprobensampling in der Nähe von LFI Punkten	14
Flächige Analyse von Daten aus sozialen Medien in Schweizer Wäldern.....	19
Evaluation von Daten aus sozialen Medien für die Erforschung von Waldnutzung und -wahrnehmung und praktische Empfehlungen	28
Wie gut ist die Verfügbarkeit und die Abdeckung von Daten aus sozialen Medien?	28
Wie repräsentativ sind Daten aus sozialen Medien?.....	29
Wie glaubwürdig sind Daten aus sozialen Medien?	30
Wer hat die Kontrolle über den Prozess und die Verfügbarkeit der Daten?.....	30
Schlussbetrachtung	32
Literaturangaben	33
Anhang A Übersicht zu Studien über Besucherverhalten und – wahrnehmung mit Datengrundlage soziale Medien	37

Zusammenfassung

Wälder bieten wichtige Erholungsfunktionen an. Um diese zu erfassen, brauchen wir deshalb Informationen zum Besuchsaufkommen, sowie zum Verhalten und der Wahrnehmung von Besucherinnen und Besuchern. Traditionelle Erhebungsmethoden wie Umfragen oder Schrittzähler sind aber aufwendig und kostenintensiv. Daher sind neue Datenquelle wie soziale Medien interessant, da sie von zahlreichen Nutzerinnen und Nutzern generiert werden, oft ohne direkten Einfluss von Forschenden. Da immer mehr Personen soziale Medien verwenden, um ihre Erlebnisse und Informationen über besuchte Orte zu teilen, steigt die Verfügbarkeit dieser Daten, was neue Möglichkeiten für die Erforschung von Erholungsnutzung mit sich bringt. Ziel dieses Berichts ist es, das Potential von Daten aus sozialen Medien für die Erforschung der Walderholungsnutzung in der Schweiz auszuloten.

Der Bericht ist gegliedert in eine Übersicht über den Stand der Forschung, eine empirische Pilotstudie anhand ausgewählter Datensätze und eine kritische Evaluation des Potentials solcher Daten.

Die Literaturanalyse zeigt, dass unterschiedliche Datenquellen verwendet werden. Diese reichen von Plattformen wie Instagram und Flickr über den Kurznachrichtendienst Twitter hin zu Daten aus persönlichen FitnessApps (GPSies, Strava) oder Wikipedia- und Wanderblog-einträgen. Themen, die anhand solcher Daten untersucht werden sind beispielsweise räumliches und zeitliches Besuchermonitoring in Schutzgebieten, Aktivitäten von Besucherinnen und Besuchern sowie die wahrgenommene Landschaft.

In der empirischen Pilotstudie untersuchen wir anhand der drei Plattformen Instagram, Flickr und Twitter die Verfügbarkeit der Daten an ausgewählten Stichprobenpunkten des Landesforstinventars (LFI). Nur sehr wenige Daten waren im unmittelbaren Umkreis der LFI Stichprobenpunkte verfügbar. Für die nachfolgenden Analysen verwenden wir deshalb sämtliche Daten innerhalb der Waldflächen. Ein Vergleich der Daten mit einem bestehenden Erholungspotentialmodell zeigte, dass die Daten aus sozialen Medien die Erholungsnutzung besonders im tourismusgeprägten, alpinen Raum höher einschätzen als das bestehende Modell, welches von einer Schätzung aufgrund der Wohnbevölkerung ausgeht. Eine Analyse des Textinhalts von sozialen Medien ergab eine Charakterisierung ausgewählter Wälder anhand von Landschaftselementen, Aktivitäten und Eigenschaften.

Aufgrund unserer Literaturanalyse und der empirischen Pilotstudie empfehlen wir, das Thema Daten aus sozialen Medien für die Erholungsnutzungsforschung weiter zu verfolgen. Kritisch zu betrachten ist dabei die Verfügbarkeit der Daten, welche bei kleinräumigen Studien oft nicht gegeben ist. Ein wichtiger Punkt ist ausserdem die Abhängigkeit von kommerziellen Firmen, welche die Plattformen betreiben. Diese Firmen können die Datenerfassung oder -bereitstellung ohne Vorankündigung unterbrechen, was für langfristiges Monitoring problematisch ist. Wir empfehlen daher, den Aufbau eigener Plattformen und Apps zur aktiven Erfassung solcher Daten zu prüfen.

Einleitung

Wälder bieten wichtige Erholungsfunktionen an. Um diese Waldfunktionen optimal zu gestalten, sind zusätzlich zu Informationen über den physischen Zustand der Wälder Informationen zum Verhalten und zur Wahrnehmung von Waldbesucherinnen und -besuchern notwendig. In der Schweiz werden detaillierte Informationen zum Zustand der Wälder über das Landesforstinventar (LFI) erhoben, und die Einstellung der Schweizer Bevölkerung zum Wald wurde anhand von Umfragen im «Waldmonitoring soziokulturell (WaMos)» untersucht. Das Ziel des Projektes WML (*Waldmonitoring soziokulturell meets Landesforstinventar*) ist es, diese beiden Erfassungsinstrumente in Bezug auf die Walderholung zu verknüpfen. Von Interesse ist dabei die Zählung von Besucherinnen und Besuchern, sowie deren Verhalten und Wahrnehmung im Bezug zur Erholungsnutzung von Wäldern.

Daten zur Besuchsfrequenz wurden bisher im Landesforstinventar (LFI) anhand von Försterbefragungen und dem Erholungsnutzungspotential abgeschätzt. Diese Methoden liefern allerdings teilweise nicht empirisch erhärtete Ergebnisse. Weitere Methoden, um solche Daten zu erheben ist die Besucherzählung im Feld, wie sie im WML Projekt (Modul 1) durchgeführt wurden. Diese Methode erwies sich als zeitaufwendig, kostspielig, und teilweise wenig verlässlich. Das Projekt WML 2 fokussiert auf die räumlich explizite Modellierung der wahrgenommenen Attraktivität von Erholungswäldern, um so die Erholungseignung von Wäldern in der Schweiz abschätzen zu können. Allerdings wird auch in WML die Besuchsfrequenz nicht wirklich empirisch erhoben. Ein weiterer Ansatz, welcher in den letzten Jahren vermehrt zur Anwendung kommt, ist die Analyse von Daten aus den sozialen Medien, welche durch die Aktivitäten von Nutzerinnen und Nutzern auf verschiedenen Plattformen wie Instagram oder Flickr anfallen. Indem Personen ihre Erfahrungen fotografieren, beschreiben und auf öffentlich zugängliche soziale Plattformen hochladen, generieren sie oft unbewusst wertvolle Daten, welche die Wissenschaft zur Beantwortung verschiedenster Fragestellungen verwendet. Solche Daten wurden beispielsweise analysiert, um Besucherhäufigkeit und -präferenzen zu untersuchen, gerade auch in Regionen, wo empirische Erhebungen fehlen oder lückenhaft sind (Sonter et al. 2016).

Im Rahmen von WML 2 gab das Bundesamt für Umwelt BAFU die vorliegende Pilotstudie in Auftrag, um das Potential der stetig wachsenden Menge an Daten aus sozialen Medien für die Erfassung von Walderholung in der Schweiz zu untersuchen.

Im Folgenden geben wir dazu als Erstes einen Überblick über soziale Medien und beleuchten in einer Literaturanalyse bisherige Forschung zu Erholungsnutzung anhand sozialer Medien. Wir zeigen dann anhand eines empirischen Beispiels in Schweizer Erholungswäldern die Datenverfügbarkeit auf drei gängigen Plattformen auf und beleuchten methodische Stärken und Schwächen. Schliesslich evaluieren wir die Möglichkeiten einer Einbindung solcher Daten ans WML, sowie an weitere Landschafts- und Waldmonitoringsinstrumente wie das LFI oder LABES und geben Empfehlungen für die Verwendung von sozialen Mediendaten für die Erholungsnutzung.

Von Tweets und Tags: Daten aus sozialen Medien für die Forschung

Der Begriff soziale Medien bezieht sich auf computerbasierte Programme, welche für den Austausch von Informationen wie Bildern, Text und Videos über digitale soziale Netzwerke genutzt werden. Von Interesse in diesem Bericht sind besonders standortbezogene Daten aus sozialen Medien, welche räumliche und zeitliche Informationen, sowie Inhalte (z.B. Fotos und Text) enthalten. Ein Beispiel solcher Daten sind mit Koordinaten und Text versehene Fotos (Abb. 1), die von Nutzerinnen und Nutzern geteilt werden.



tags: *bärenhütte, fürstenwald, chur, graubünden, wald, bündner wald, grison, switzerland, forest, nature, GRischashot*

Abb 1. Mit Koordinaten und tags versehenes Bild auf Flickr von Roman Schurte
(öffentliche Domäne, creative commons Lizenz CC0 1)

So hat der Nutzer R. Schurte in einem Bündner Wald ein Foto gemacht, es mit Koordinaten versehen und mit verschiedenen Tags beschrieben (z.B. *Bärenhütte, Chur, Bündner Wald*), um es auf die Plattform «Flickr¹» hochzuladen. Nebst Flickr existiert eine Reihe weiterer Plattformen, über die Nutzerinnen und Nutzer Inhalte teilen, wie zum Beispiel «Instagram»² und der Kurznachrichtendienst «Twitter»³. Die Plattformen unterscheiden sich darin, welche Inhalte geteilt werden. Instagram und Flickr werden vor allem benutzt, um Fotos zu teilen, welche mit Kurzbegriffen versehen sind (Di Minin, Tenkanen, and Toivonen 2015). Diese sogenannten «hashtags» (Instagram) oder «tags» (Flickr), wie zum Beispiel «landscape», helfen anderen Nutzerinnen und Nutzern, Inhalte zu finden und diese wiederum zu teilen. Die Urheberinnen und Urheber der Inhalte haben ein Interesse daran, diese Fotos zu «taggen», denn dadurch werden sie erst für andere auffindbar, und können mit Kommentaren und «likes» versehen werden, was wiederum das soziale Ansehen der Urheberinnen und Urheber erhöht.

Bei zahlreichen Plattformen sind die Daten für die Öffentlichkeit und über sogenannte Programmierschnittstellen (Application Programming Interfaces oder kurz APIs) zugänglich.

¹ www.flickr.com

² www.instagram.com

³ www.twitter.com

Über die API können Daten sowohl nach Stichwörtern oder für einen bestimmten Raumausschnitt gesucht werden.

Die Kombination von grafischen Inhalten mit Textdaten und Koordinaten, sowie die Zugänglichkeit der Information machen Daten aus sozialen Medien interessant für die Erforschung von Erholungsnutzung. Im Folgenden geben wir einen Überblick zur bestehenden Literatur in diesem Bereich.

Literaturreview zu sozialen Medien für die Erforschung von Erholungsnutzung

Da sich nur wenige Studien spezifisch mit dem Besuchermonitoring in Erholungswäldern beschäftigen, haben wir den Literaturreview auf Studien ausgeweitet, welche soziale Medien als Grundlage verwenden, um Aussagen über die Erholungsnutzung und Wahrnehmung von Erholungslandschaften zu treffen. Dies schliesst somit Studien aus, welche soziale Medien zur Überwachung von Ökosystemen verwendet haben, wie beispielsweise das Monitoring von Schädlingen (Daume, Albert, and von Gadow 2014). Zur Verständlichkeit gliedern wir unseren Überblick der Literatur in drei Themenbereiche: «Besucherhäufigkeit», «Besucherverhalten», sowie «Präferenzen und Wahrnehmung», wobei gewisse der erwähnten Studien mehrere dieser Themenbereiche abdecken können. Sämtliche für den Literaturreview verwendeten Studien sind in tabellarischer Ansicht und mit einem Kurzüberblick zum Inhalt in Anhang A zusammengefasst.

Besucherhäufigkeit

Eine wichtige Entscheidungshilfe für das Management von Erholungsgebieten ist die Besucherhäufigkeit. Durch eine Abschätzung der Anzahl NutzerInnen und deren Verteilung in einem Erholungsgebiet können negative Effekte frühzeitig identifiziert und falls notwendig Massnahmen zur Besucherlenkung eingeleitet werden (Orsi and Geneletti 2013). Zudem kann eruiert werden, wie Faktoren, die durch das Management beeinflusst werden (z.B. Wegnetz) sich auf die Besucherhäufigkeit auswirken. Traditionelle Messmethoden sind mechanische Wegzähler oder Umfragen vor Ort, die sehr aufwendig und daher kostspielig sind (English et al. 2002). Um das Problem der kostspieligen Datenaufnahme zu beheben, werden zur Abschätzung der Besucherfrequenz in Schutzgebieten zunehmend Daten aus den sozialen Medien herbeigezogen (Wood et al. 2013; Norman and Pickering 2017; Tenkanen et al. 2017). Anhand einiger konkreter Beispiele empirischer Studien zeigen wir nachfolgend die Verwendung von Daten aus sozialen Medien für das Besuchsmonitoring auf.

Studienbeispiele. In einer Studie zu über 800 Freizeitattraktionen weltweit (darunter Nationalparks in Uganda, Strände in Kalifornien und Disneyland) zeigte sich ein verlässlicher, statistisch signifikanter Zusammenhang zwischen der Anzahl an gezählten Besucherinnen und Besuchern und einer aus Flickr-Fotos berechneten Messgrösse (Wood et al. 2013). Diese Grösse wird anhand georeferenzierter Flickr-Fotos als Anzahl Tage errechnet, die eine Nutzerin oder ein Nutzer in einem Gebiet verbracht hat. Dies sind also Tage, an denen eine Person an einem Standort mindestens ein Foto auf Flickr hochgeladen hat. Allerdings fanden die Forschenden, dass sich die Übereinstimmung von *in situ* Zählungen und Flickr-Daten zwischen verschiedenen Freizeitattraktionen stark unterscheidet und empfehlen daher, nur relative Veränderungen anhand von Flickrdaten zu betrachten (Wood et al. 2013). In einer weiteren Studie verwendeten Forschende georeferenzierte Flickr-Fotos, um die Besuchshäufigkeit in Schutzgebieten in Vermont (USA) abzuschätzen. Sie fanden dabei eine signifikante Korrelation

mit *in situ* Zählungen. Anhand der Korrelation konnte eine Abschätzung der Gesamtbesucherzahl für einen Zeitraum über mehrere Jahre errechnet werden, in denen keine Zählungen an Ort und Stelle vorgenommen worden waren. Allerdings zeigte das Modell eine geringe Vorhersagekraft ($R^2 = 0.22$). Die AutorInnen führen dies auf die hohe Variabilität in den Aktivitäten und der Demographie von Erholungssuchenden zurück, die sich auf die Besuchsfrequenz auswirkt. Andere Studien zeigten jedoch, dass anhand von sozialen Medien auch erheblich höhere R^2 erreicht werden können. Im meistbesuchten Nationalpark Finnlands fanden Forschende beispielsweise eine hohe signifikante Korrelation zwischen Besuchszahlen und Instagram-users ($R^2 = 0.67$, $p\text{-value} < 0.05$), (Heikinheimo et al. 2017), siehe auch Abb. 2.

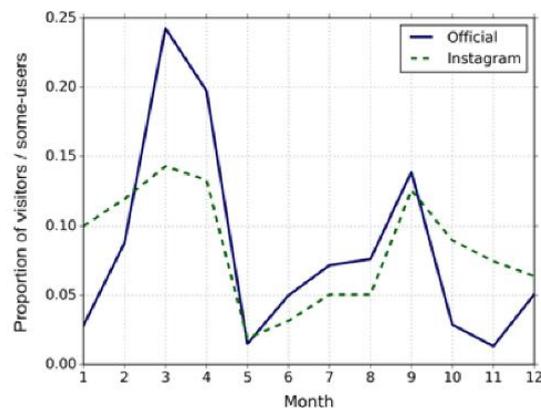


Abb. 2 Besuchszahlen in einem finnischen Nationalpark gemäss Instagram-users (gestrichelte Linie) und Umfragen (durchgezogene Linie), (Heikinheimo et al. 2017)

Ähnliche Zahlen wurden für Besucher von Seen in Iowa und Minnesota anhand von Flickr ($R^2 = 0.65$) (Keeler et al. 2015) oder von urbanen Grünräumen in Minnesota anhand von Flickr ($R^2 = 0.67$) und Twitter Daten ($R^2 = 0.64$) erreicht (Donahue et al. 2018). Zudem waren im letztgenannten Beispiel die anhand der zwei unterschiedlichen Plattformen errechneten Besuchszahlen stark miteinander korreliert ($R^2 = 0.90$), was darauf hindeutet, dass unterschiedliche Plattformen ein ähnliches Abbild der Besuchenden zeichnen (Donahue et al. 2018). Eine weitere Studie aus Nationalparks in den USA stellte fest, dass der Zusammenhang zwischen Fotos auf Flickr und tatsächlicher Anzahl BesucherInnen bei hochfrequentierten Parks grösser ist als bei wenig frequentierten Parks (Sessions et al. 2016). Die AutorInnen empfehlen daher, Besuchsfrequenzschätzungen anhand sozialer Medien nur in stark besuchten Parks anzuwenden (z.B. Yellowstone, Grand Canyon etc.)

Auch Daten aus FitnessApps werden zur Abschätzung von Besuchszahlen verwendet. Beispielsweise korrelierte die Anzahl Nutzerinnen und Nutzern von FitnessApps (MapMyFitness und GPSies) in Australischen Schutzgebieten stark mit den tatsächlich gemessenen Besucherströmen ($R^2 = 0.80$) (Norman and Pickering 2017). Allerdings waren die Zählungen in verschiedenen Grössenordnungen. So kommen auf 150 NutzerInnen von MapMyFitness rund 8000 tatsächlich gezählte Personen auf demselben Wanderweg.

Durch die sich ständig verändernden Plattformen erhalten Forschende stets neue Möglichkeiten, Daten zu erhalten. So wurden auch OpenStreetMap (OSM) oder Wikipedia-Artikel für das Besuchermonitoring verwendet, indem die Anzahl Veränderungen an den Daten (Kartenelemente oder Artikel) als Indikator für die Besucherfrequenz gezählt wurde (Levin,

Lechner, and Brown 2017). In australischen Parks korrelierte diese Zahl der Editierungen von Wikipedia-Artikeln und OSM-Geodaten sowohl mit der Anzahl Flickrfotos als auch mit offiziellen Besucherstatistiken (Levin, Lechner, and Brown 2017).

Auch für die Analyse zeitlicher Trends bieten Daten aus sozialen Medien neue Möglichkeiten der rückwärtsblickenden Analyse. Es können zeitliche Veränderungen in Besuchszahlen eruiert werden, und starke Schwankungen identifiziert werden, die sich zum Beispiel auf spezifische Ereignisse beziehen, wie Festivals oder saisonale Trends: Die alljährlichen Herbst-Besucherströme in Vermont zeichnen sich in Flickrdaten ab (Wood et al. 2013), und in der italienischen Cilento Region lassen sich saisonale Unterschiede in Besucherströmen anhand von Tweets nachvollziehen (Chua et al. 2016).

Erkenntnisse:

- ➔ Relative Veränderungen der Besucherhäufigkeit können anhand von sozialen Medien in Regionen mit viel Daten abgeschätzt werden.
- ➔ Für Veränderungen in absoluten Besucherzahlen wird die Kombination von Daten aus sozialen Medien oft mit *in situ* Zählungen verbunden, um die mit Unsicherheit behafteten Daten zu validieren und so ein genaueres Gesamtbild zu erhalten

Besucherverhalten

Die von Besucherinnen und Besuchern in Erholungsräumen durchgeführten Aktivitäten und deren Lokalisierung sind wichtige Informationen für das Management von Erholungsgebieten. Während traditionelle Erhebungsmethoden auf Interviews setzen, um die tatsächlich durchgeführten oder erwünschten Aktivitäten abzufragen, gibt es Bestrebungen, solche Erkenntnisse anhand von sozialen Medien ohne direkten Kontakt mit den Besuchenden zu eruieren. Im Gegensatz zu Umfragen, deren Ergebnisse durch die Interaktion mit Forschenden entstehen, fallen Daten aus sozialen Medien von Erholungssuchenden an, ohne dass sie mit Forschenden interagieren. Da NutzerInnen sozialer Medien die Position ihrer Fotos mehrheitlich sehr zuverlässig angeben (Hollenstein and Purves 2010) lassen sich dadurch Informationen zu ihren Aktivitäten gewinnen, die eher Einblicke in tatsächliches Verhalten erlauben, obwohl wir auch hier von einer Beeinflussung der sozialen Erwünschbarkeit ausgehen müssen (z.B. lassen sich nicht alle Aktivitäten gleich gut oder wirksam auf sozialen Medien «inszenieren»).

Studienbeispiele. Anhand einer Inhaltsanalyse von Instagramfotos in einem finnischen Nationalpark ermittelten Forschende beispielsweise Wandern und Langlaufen als die am häufigsten fotografierten Aktivitäten. Diese wurden in einer Umfrage unter Parkbesucherinnen und -besuchern ebenfalls am häufigsten genannt (Heikinheimo et al. 2017). Allerdings gab es auch Unterschiede zwischen den Methoden. Aktivitäten wie Snowboarden und Tourenskifahren wurde häufiger auf Instagramfotos abgebildet als in Umfragen erwähnt, während Picknicken und Beerenpflücken öfter in Umfragen genannt wurde (Heikinheimo et al. 2017). Gewisse Aktivitäten wurden nur in Instagramfotos gefunden und waren nicht in der Umfrage genannt worden, so zum Beispiel wie Husky- und Rentierschlittenfahren, Kiteboarding

und Schneemobilfahren. Diese Unterschiede sind vermutlich darauf zurückzuführen, dass eher ältere Leute in Umfragen mitmachen, während Instagram eher von jüngeren Personen verwendet wird. Instagram bietet sich zudem an, über eher spektakuläre Aktivitäten zu berichten, um so viele «likes» zu erhalten. Wir erwarten beispielsweise tendenziell mehr «likes» für ein Foto einer Huskyschlittentour als für Beerensammeln oder Picknicken. Die Methoden bilden ausserdem auch unterschiedliche demographische Stichproben ab, die sich in ihren Aktivitäten unterscheiden. Auch konnten gut zeitliche Unterschiede in den Aktivitäten festgestellt werden: Wandern ist stärker im Sommer vertreten und im Winter dominieren Schneesportarten – was sich wiederum mit Erkenntnissen aus den Umfragen deckt (Heikinheimo et al. 2017).

Eine weitere Quelle für Daten aus sozialen Medien sind Fitness-Apps, bei denen Nutzerinnen und Nutzer ihre zurückgelegten Streckenprofile für verschiedene Aktivitäten auf Plattformen hochladen, um sich mit anderen zu vergleichen oder Routenvorschläge zu machen. Daten von MapMyFitness und GPSies wurden beispielsweise in Schutzgebieten in Australien verwendet, um die relative Popularität von Velofahren und Wandern in unterschiedlichen Parks zu untersuchen (Norman and Pickering 2017). Die Forschenden konnten zudem aufzeigen, welche Trails bei den Besucherinnen und Besuchern beliebter waren als andere. Die Daten offenbarten ausserdem illegale Verhaltensweisen, da gewisse Nutzerinnen und Nutzer Streckenprofile auf einen Berg hochgeladen hatten, der als heilige Stätte unter strengstem Schutz steht und dessen Betreten absolut verboten ist (Norman and Pickering 2017). Eine Studie im Sintra-Cascais Naturpark analysierte die Nutzungsintensität von Mountainbikern anhand von nutzergenerierten Daten der Plattformen GPSies und Wikiloc.com. Ein Vergleich der beiden Plattformen zeigte ein ähnliches Bild der Nutzung von Mountainbiketrails auf. Die Studie zeigte ebenso illegales Verhalten von Bikern auf, die sich abseits offizieller Pfade bewegten (Campelo and Nogueira Mendes 2016). Die Analyse von tracking-Daten aus Fitness-Apps wirft daher aber auch ethische Fragen auf, die in der Literatur kontrovers diskutiert werden. Einerseits wird argumentiert, dass eine Analyse solcher Daten problemlos möglich sei, da die Daten von NutzerInnen frei zur Verfügung gestellt werden (Campelo and Nogueira Mendes 2016). Andererseits wird argumentiert, dass auf die Forschung mit solchen Daten in streng geschützten Gebieten verzichtet werden sollte, da für die NutzerInnen negative Konsequenzen entstehen können, wenn die auf Fitnessapps hochgeladenen tracks illegale Nutzung aufzeigen (Korpilo, Virtanen, and Lehvavirta 2017).

Eine Studie in Finnland setzte daher auf freiwillige Parknutzerinnen und -nutzer, die ihre GPS-tracks von verschiedenen Fitness-Apps (Polar, Endomondo, Strava) der Forschung zur Verfügung stellten und explizit eine formelle Einverständniserklärung ausfüllten (Korpilo, Virtanen, and Lehvavirta 2017). Anhand dieser «gespendeten» tracks analysierten die Forschenden Nutzungsintensität im Stadtpark und Erholungswald Keskuspuisto in Helsinki. Sie fanden dabei Hot- und Coldspots in den Daten, welche gut mit Felddaten zur Nutzung übereinstimmten. So trat beispielsweise zerstörte Vegetation und Erosion in Hotspots der social media Daten auf, nicht aber in Coldspots (Korpilo, Virtanen, and Lehvavirta 2017). Forschende eruierten zudem Hotspots von outdoor-Erholungsaktivitäten in ganz Europa, indem sie Flickr und Instagram Daten anhand tags in fünf verschiedenen europäischen Sprachen auswählten, die mit Erholungsaktivitäten in Verbindung stehen (z.B. *hiking*, *wandern*, *pasear*, etc.) und deren räumliche Verteilung analysierten (van Zanten et al. 2016).

Erkenntnisse:

- ➔ Besuchsverhalten, wie beispielsweise in Erholungsgebieten ausgeübte Aktivitäten, können anhand von Daten aus sozialen Medien erhoben werden
- ➔ Methodisch werden qualitative Inhaltsanalysen von Bildern durchgeführt, oder es werden georeferenzierte Fotos, GPS-tracks (z.B. von Fitness-Apps) oder Textdaten (tags wie z.B. «wandern», «hiking») verwendet

Präferenzen und Wahrnehmung

Daten aus sozialen Medien stellen eine neue Art von Informationsquellen dar, welche nicht wie bei einer Umfrage durch die Präsenz von Forschenden im Feld beeinflusst werden. Soziale Medien stellen somit eher beobachtete Präferenzen (sogenannte «revealed preferences») und tatsächliches Verhalten dar. Eine andere Art der Präferenzen sind «stated preferences», also angegebene Präferenzen, wie sie zum Beispiel anhand von Umfragen oder öffentlichen Anhörungen von der Bevölkerung eingeholt werden können (Wardman 1988). Allerdings sind gerade an Anhörungen nicht alle Bevölkerungsgruppen repräsentiert und Themen können stark umstritten diskutiert werden, und sich auf negative Voten fokussieren (Moote, McClaran, and Chickering 1997). Im Gegensatz dazu sind Daten aus sozialen Medien nicht auf ein einzelnes (negatives) Thema fokussiert und die geäußerten Meinungen weniger kontrovers, da der Zweck der sozialen Medien oftmals das Reflektieren und Teilen der eigenen Erlebnisse ist (Barry 2014). Daten aus sozialen Medien können deshalb als Ergänzung zu traditionellen Erhebungsmethoden verwendet werden, um Meinungen und Einstellungen zu gewissen Themen in Schutzgebieten oder Erholungszone zu eruieren, oder um Landschaftswahrnehmung räumlich zu untersuchen.

Studienbeispiele. In Kalifornien analysierte eine Forscherin anhand georeferenzierter Flickr-Daten mit den *tags* «cow» und «grazing» und qualitativer Inhaltsanalysen von Kommentaren und Beschreibungen die öffentliche Einstellung zu Kühen und der Beweidung auf öffentlichem Grund zur Prävention von Waldbränden (Barry 2014). Eine weitere Anwendung von Daten aus sozialen Medien ist die Kartierung von kulturellen Ökosystemdienstleistungen, wie beispielsweise dem ästhetischen Wert einer Landschaft (Oteros-Rozas et al. 2017; Figueroa-Alfaro and Tang 2017; Guerrero et al. 2016). Als Proxy für den ästhetischen Wert wird oft die Anzahl NutzerInnen verwendet, die ein einem bestimmten Gebiet Fotos auf eine Plattform hochladen. Der dadurch errechnete Wert kann mit anderen Kenngrößen in Verbindung gebracht werden. So fanden Forschende heraus, dass in Cornwall die Anzahl UserInnen, welche Fotos hochladen, in Küstenregionen höher ist als im Inland, und generell negativ mit der landwirtschaftlichen Wertschöpfung als Indikator für Intensität der Nutzung und der Bevölkerungsdichte korreliert ist (Casalegno et al. 2013). Die Anzahl hochgeladener Fotos auf Flickr, Instagram und Panoramio, die zuvor nach tags mit outdoor-Aktivitäten gefiltert worden waren, wurden zudem verwendet, um den kulturellen Wert von Landschaften in Europa zu bestimmen (van Zanten et al. 2016). Dabei zeigte sich, dass gewisse hochgradig bekannte Attraktionen, wie der «Camino de Santiago» die Anzahl hochgeladener Bilder stark beeinflusst. Die Anzahl hochgeladener Bilder auf Panoramio wurde ausserdem als Teil eines Indikators für charakteristische Kulturlandschaften in Europa verwendet (Tieskens et al. 2017).

Anhand von Textdaten aus sozialen Medien lassen sich zudem Beschreibungen der Landschaftsqualität extrahieren, wodurch sich verschiedene Erholungslandschaften miteinander vergleichen lassen. Eine Studie in der Schweiz verglich Flickr-tags aus zehn Erholungslandschaften mit Beschreibungen aus Wanderblogs und Interviews (Wartmann, Acheson, and Purves 2018). Vergleiche der verschiedenen Beschreibungen von Erholungslandschaften zeigte, dass sich Landschaften aufgrund der genannten Landschaftsbegriffe unterscheiden lassen. Aus sozialen Medien extrahierte Textinformationen haben deshalb Potential, um für Landschaftscharakterisierungen verwendet zu werden (Wartmann, Acheson, and Purves 2018; Wartmann et al. 2015).

Erkenntnisse:

- ➔ Daten aus sozialen Medien bieten die Möglichkeit, das dargestellte Verhalten und die von Nutzerinnen und Nutzern angegebenen Präferenzen zu untersuchen.
- ➔ Methodisch werden Anzahl Bilder als Indikatoren für wahrgenommene Schönheit verwendet oder um den wahrgenommenen Wert von Erholungslandschaften abzuschätzen
- ➔ Textinhalte von Wanderblogs und tags werden analysiert, um die Wahrnehmung von Landschaften und deren Charakter aufgrund von Sprache zu untersuchen

Testweise Analyse von Daten aus sozialen Medien zur Erforschung von Erholungswäldern in der Schweiz

Wie die Literaturübersicht gezeigt hat, finden sich zahlreiche Beispiele der Erforschung von Besuchernutzungsverhalten, -intensität und Landschaftswahrnehmung anhand von Daten aus sozialen Medien. Da diese Studien jeweils in spezifischen geographischen, soziokulturellen und ökologischen Kontexten durchgeführt wurden, sind deren Ergebnisse nicht direkt auf die Thematik von Erholungswäldern in der Schweiz übertragbar. Die betrachteten Studien, welche erfolgreich statistische Zusammenhänge zwischen Bilddichte auf sozialen Medien und Besucherströmen nachweisen konnten, beziehen sich zum Beispiel auf hochfrequentierte Pärke mit einer internationalen Ausstrahlung, wie der Yellowstone Nationalpark, welcher jährlich Millionen von Besucherinnen und Besuchern anzieht (Wood et al. 2013), oder hochfrequentierte städtische Pärke und Stadtwälder (Korpilo, Virtanen, and Lehvavirta 2017). Ob die Methodik auf weniger stark frequentierte Gebiete mit eher lokaler Nutzung übertragbar ist, bleibt deshalb zu testen.

Wo die Landschaftswahrnehmung im Vordergrund steht, sind Anzahl Bilder aus sozialen Medien ein beliebter Proxy für die Landschaftsschönheit geworden. Allerdings zeigen Studien, dass die Anzahl Bilder stark durch die Erreichbarkeit von Orten geprägt ist, und nicht nur durch deren landschaftliche Schönheit (Tenerelli, Demšar, and Luque 2016). Deshalb bieten sich möglicherweise inhaltliche Analysen von Daten aus sozialen Medien an, die über das reine Zählen von Daten hinausgehen, um die Wahrnehmung und Nutzung von Erholungswäldern in der Schweiz zu untersuchen.

Im Folgenden stellen wir eine testweise Analyse von Daten aus sozialen Medien vor. Das Ziel ist, anhand dieser empirischen Pilotstudie das Potential von sozialen Mediendaten zur Erforschung von Erholungswäldern in der Schweiz abzuschätzen. Wir analysieren hier zwei verschiedene Stichprobenkonzepte: a) die Punktstichprobe aus dem LFI und b) die flächige Analyse von Waldgebieten.

Der Fokus dieser empirischen Pilotstudie liegt einerseits auf der Überprüfung der Datenverfügbarkeit in Wäldern auf verschiedenen Plattformen. Andererseits setzen wir uns exemplarisch mit den Inhalten auseinander und verbinden diese mit bestehenden Modellen. So vergleichen wir das Modell des Erholungspotentials aus dem Landesforstinventar mit Daten aus den sozialen Medien. Zuletzt untersuchen wir für ausgewählte Wälder Textinhalte aus sozialen Medien, um über die verwendete Sprache und Semantik einen Einblick in die Wahrnehmung dieser Waldgebiete zu erhalten.

Verfügbarkeit von Daten aus sozialen Medien in Schweizer Wäldern

Um überhaupt soziale Medien für Analysen und schliesslich als mögliche Indikatoren zu verwenden, muss erst deren Verfügbarkeit gegeben sein. Wir überprüfen deshalb testweise für eine kleine Stichprobe von Standorten die Datenverfügbarkeit. Im Projekt «WaMos-meets LFI» (WMS) ist die Verknüpfung des standardisierten Landesforstinventars mit dem soziokulturellen Waldmonitoring zentral. Die Stichprobenpunkte für das WML (Modul 1) wurden ausgewählt als

LFI Stichprobenpunkte, die in gut erreichbaren Wälder in Stadtnähe mit hohem Nutzungspotential gelegen waren. Aus diesen Standorten wählen wir für unsere Analyse zehn Punkte aus, die sich in der Nähe von LFI Stichprobenpunkten befinden. Davon sind drei Punkte im WML Modul 1 befragt worden (Dolder, Aarau, Ebmatingen), weitere vier werden in den WML Modulen 3 bis 5 befragt (Locarno, Arosa, Scuol, S-chanf), und drei sind allein LFI-Standorte (Uetliberg, Neuchâtel, Ovronnaz). Als zweite Analyse wählen wir die Polygone des LFI-Datensatzes, um für die gesamte Waldfläche die Verfügbarkeit von Daten aus sozialen Medien zu überprüfen.

Stichprobensampling in der Nähe von LFI Punkten

An zehn mit dem LFI/WML verbundenen Stichprobenpunkten sammeln wir im Umkreis von 5 Kilometern Daten aus den Jahren 2017 und 2018. Wir fokussieren hier auf die drei populären sozialen Netzwerke Flickr, Twitter und Instagram. Für diese Pilotstudie verwenden wir nur bereits georeferenzierte Inhalte und gehen nicht auf Verfahren der Georeferenzierung ein, anhand derer nicht-georeferenzierte Inhalte aufgrund von Ortsnamen mit Koordinaten versehen werden könnten (Hill 2009).

Insgesamt wurden an den zehn Standorten insgesamt 214'981 Datenpunkte gesammelt. Instagram wies dabei die grösste und Flickr die kleinste Menge an Daten auf:

- 2'590 Datenpunkte aus «Flickr» (Zeitraum: 9.Okt 2017 - 26. Jan 2018)
- 4'106 Datenpunkte aus «Twitter» (Zeiträume⁴: 9. Okt 2017 - 27. Okt 2017; 31. Okt 2017 - 6. Nov 2017; 22. Nov. 2017 - 7. Jan 2018)
- 208'285 Datenpunkte aus «Instagram» (Zeiträume¹: 6. Okt 2017 - 5.Nov 2017; 22. Nov 2017 - 25. Jan 2018)

Typisch für solche Datensätze ist die ungleichmässige Verteilung der Daten zwischen Standorten, aber auch zwischen Plattformen (Abbildung 3).

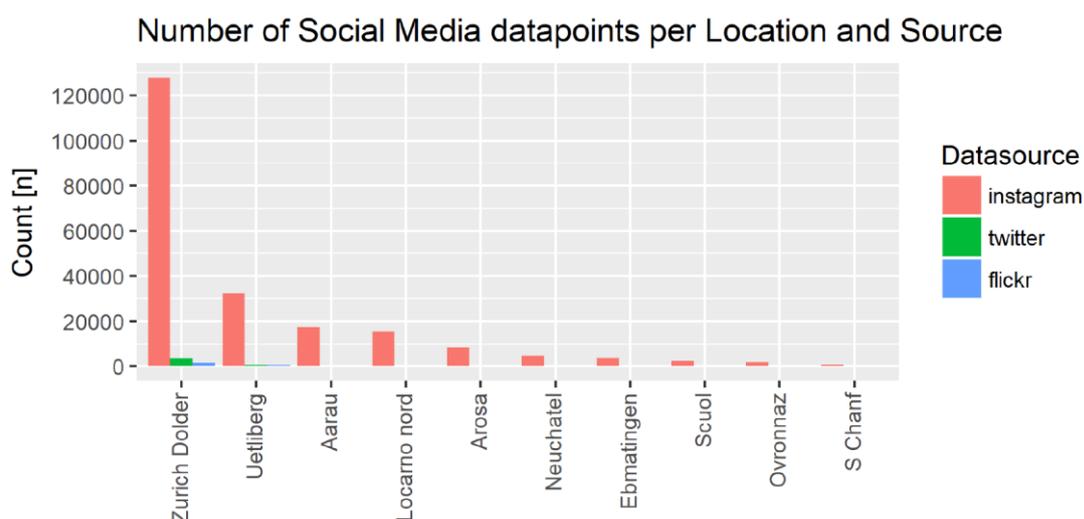


Abb. 3 Säulendiagramm der Anzahl Punkte aufgeteilt nach Plattformen

⁴ Die Unterbrüche in den Zeiträumen sind auf Serverfehler zurückzuführen

Um die Datenquellen untereinander zu vergleichen, untersuchten wir die statistische Korrelation der Anzahl Punkte pro Standort zwischen den Datenquellen. Es korrelierten dabei alle Datensätze untereinander (Twitter und Instagram mit ρ : 0.964; p -wert < 0.01; Instagram und Flickr mit ρ : 0.818; p -wert < 0.01; Flickr und Twitter mit ρ : 0.855; p -wert < 0.01). Wo in einem Datensatz viele Punkte vorhanden sind, ist dies also auch in anderen Datensätzen der Fall. Dies bedeutet, dass auch Datensätze mit einer begrenzten Anzahl an Datenpunkten verwendet werden könnten, um relevante Resultate zu erzielen, da die Häufigkeit der Daten in allen Datensätzen ähnlich verteilt ist.

Die Daten zeigen eine starke räumliche Clusterbildung. Orte mit hoher Datenkonzentration befinden sich grösstenteils an einfach zu erreichenden, meist urbanen Standorten (Abb. 4).

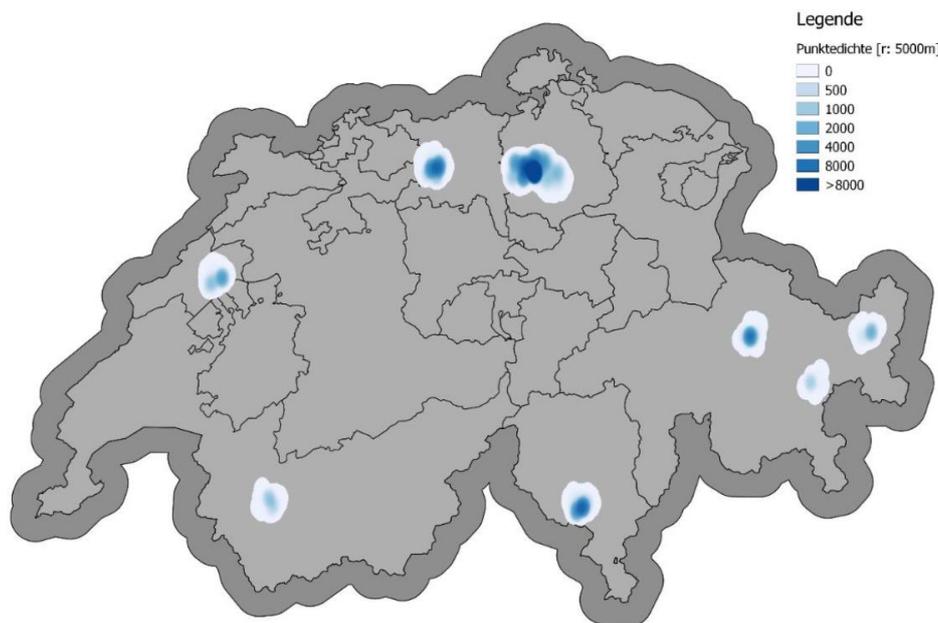


Abb. 4 Anzahl Punkte im Umkreis von 5km an den 10 ausgewählten LFI-Stichprobenpunkten

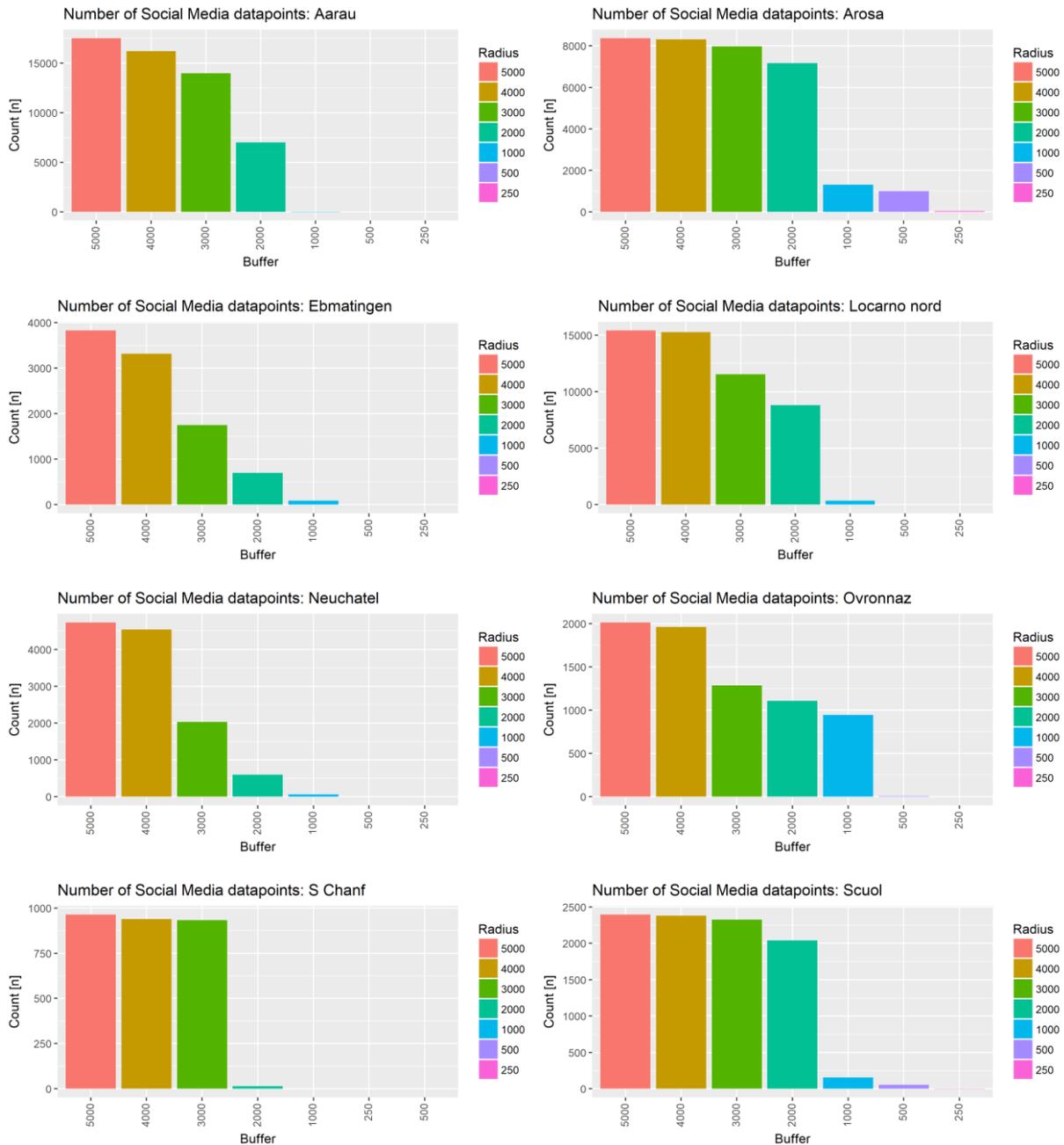
Da Daten in einem Umkreis von 5 Kilometern miteinbezogen wurden, bedeutet dies, dass je weiter die Daten von einem Standort entfernt sind, desto weniger aussagekräftig sie für diesen Standort sein werden.



Abb. 5 Anzahl Punkte im Umkreis von 5km an den 10 ausgewählten Stichprobenpunkten

Um die räumliche Verteilung der Daten um die Stichprobenpunkte genauer zu analysieren, berechneten wir die Anzahl Punkte in verschiedenen Radien. Dies ist beispielhaft illustriert für Neuchâtel (Abb. 5). In Abbildung 5 wird ersichtlich, dass sich der Grossteil der Daten nicht im Umkreis um den Stichprobenpunkt befindet, sondern im weiteren Umkreis.

Auch an den anderen Standorten zeigt sich ein ähnliches Bild, was wir in den folgenden Säulendiagrammen aufzeigen: Der Grossteil der Datenpunkte befindet sich im Umkreis von zwei bis fünf Kilometern vom Stichprobenpunkt entfernt und nur wenige Fotos wurden im unmittelbaren Umkreis von bis zu 500 Metern hochgeladen (Abb. 6).



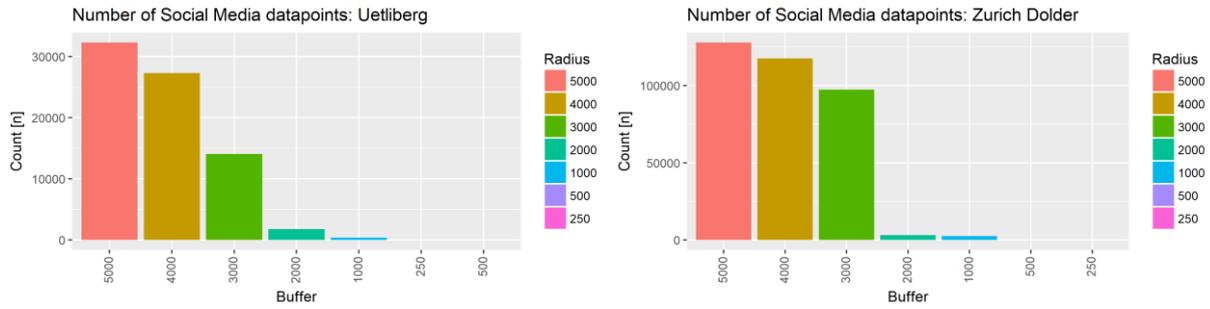


Abb. 6 Anzahl Punkte in Umkreisen von 250m bis 5km an den 10 ausgewählten Stichprobenpunkten

Eine Aufschlüsselung in die drei verschiedenen Plattformen zeigt eine ähnliche Verteilung der Daten innerhalb der Radien für die unterschiedlichen Plattformen (Abb. 7).

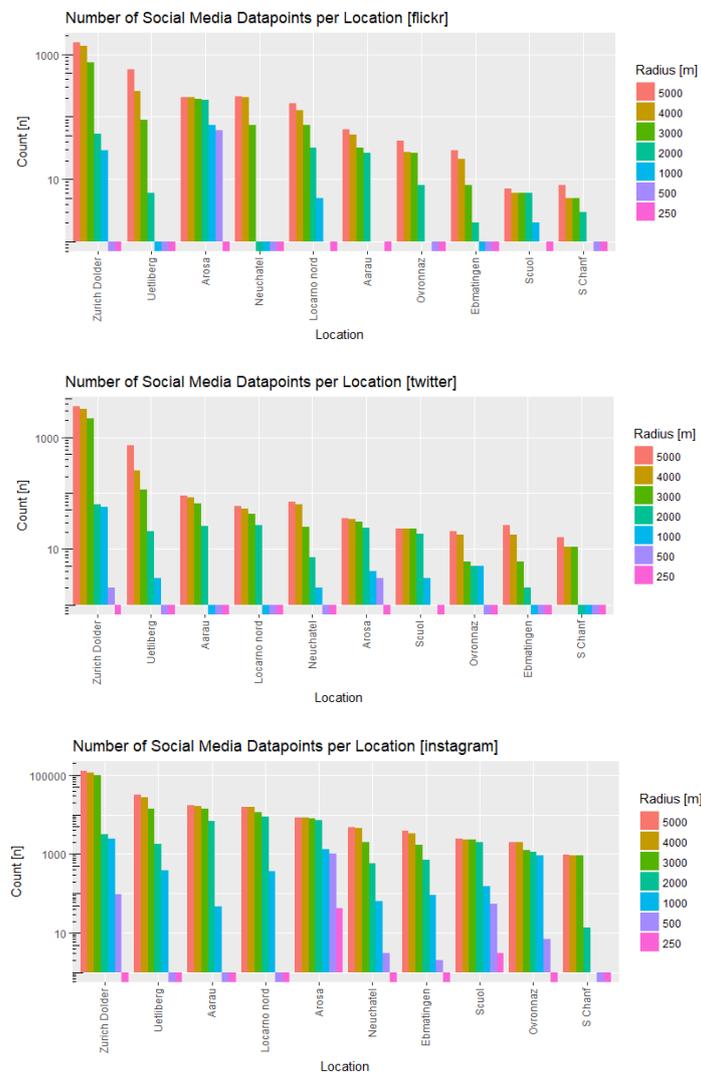


Abb. 7 Anzahl Datenpunkte in Flickr, Twitter und Instagram in verschiedenen Radien um die Stichprobenpunkte

Weiter zeigt die Abbildung 7 auch, dass es teilweise Unterschiede zwischen den einzelnen Standorten und Plattformen gibt. In Flickr finden wir um den Standort Arosa beispielsweise eine grosse Anzahl Punkte im Umkreis von 1000-2000m. Anders beim Uetliberg, wo die meisten Punkte im Abstand von 4000m – 5000m zu finden sind. Interessant ist auch, dass Arosa im

Flickr-Datensatz die dritthöchste Anzahl Punkte vorweist, im Instagram-Datensatz hingegen nur die fünfhöchste und im Twitter-Datensatz die sechshöchste Anzahl Punkte.

Um die Anzahl Punkte in Abhängigkeit zur Distanz zu visualisieren, erstellten wir Diagramme, welche die kumulierte Anzahl Punkte als Funktion der Distanz zu den Standorten zeigen (Abb. 8: Beispiel Neuchâtel und Arosa). Während sich beim Standort in der Nähe von Neuchâtel praktisch keine Punkte von Flickr im Umkreis von bis zu 2 Kilometern findet, steigt die Anzahl Punkte bei einem Radius ab 3 Kilometern rapide an, auf Instagram ist jedoch ein flacherer Anstieg zu verzeichnen.

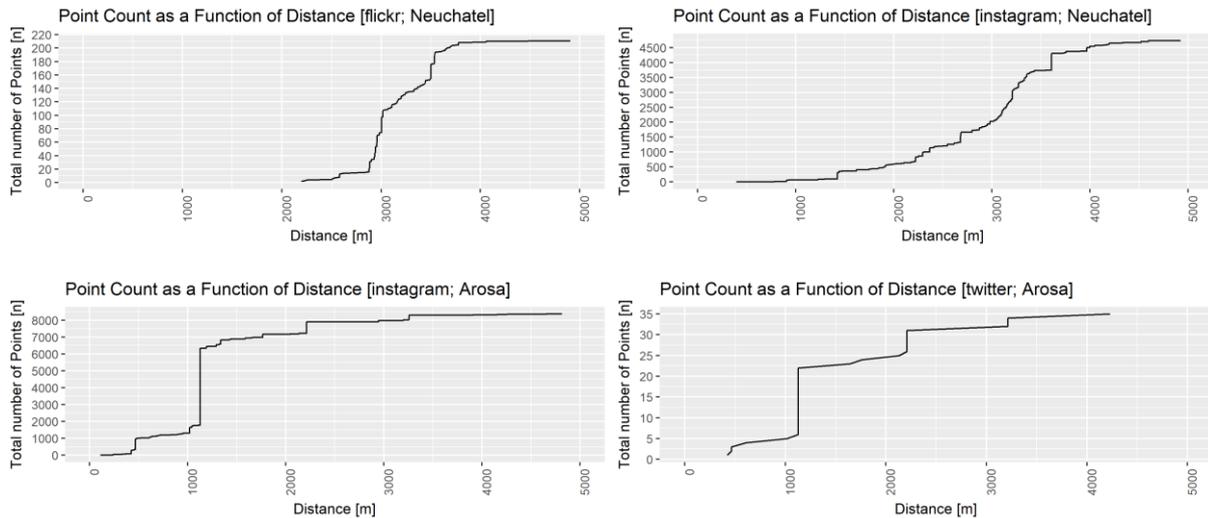


Abb. 8 Kumulative Anzahl Datenpunkte um Stichprobenpunkte Neuchâtel und Arosa

Bei den Daten von Instagram vom Standort in der Nähe von Arosa zeigt sich ein stufenhafter Anstieg beim Radius 1 Kilometer sowohl auf Instagram und danach ein flacherer Anstieg, während bei Twitter der Anstieg stufenhaft bleibt. Da wir gesehen haben, dass sich relativ wenig Daten im engen Umkreis um die vorgegeben Stichprobenpunkte finden, stellt sich die Frage: Wie verhält sich nun die Anzahl georeferenzierter Fotos im Bezug zum Untersuchungsgegenstand Wald? Das Beispiel vom Uetliberg (Abb. 9) zeigt klar: Der Grossteil der Datenpunkte befindet sich nicht innerhalb der Waldfläche, sondern in den umliegenden städtischen Gebieten im weiteren Umkreis.

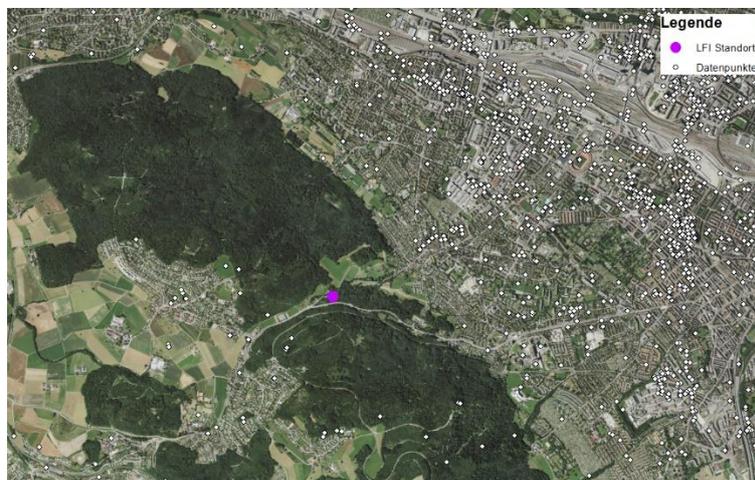


Abb. 9 Georeferenzierte Bilder auf Flickr, Twitter und Instagram am Uetliberg

Die Verfügbarkeit der Daten scheint zudem stark mit Verkehrswegen in Verbindung zu stehen. Daten kommen fast ausschliesslich an sehr einfach zu erreichenden Standorten vor, und nicht innerhalb von Waldstücken, die nicht über breite Fusswege erschlossen sind. Dies deckt sich mit persönlichen Beobachtungen innerhalb der Wälder. Obschon auch in abgelegenen Waldstücken ab und zu BesucherInnen vorbeikommen, ist die Besuchshäufigkeit gesamthaft gesehen sehr tief. Solche tiefen Frequenzen lassen sich schwer mit Daten aus sozialen Medien erfassen, die deshalb eher geeignet sind, um hohe Nutzungsintensitäten anhand Daten vieler verschiedener NutzerInnen zu erfassen.

Die Auswertungen der Datenverfügbarkeit um ausgewählte Stichprobenpunkte zeigen, dass Daten aus den sozialen Medien vor allem in urbanen Zentren und an besonderen Orten (z.B. Dolder in Zürich oder Arosa als touristische Destination) vorkommen. Die Daten zeigen somit nur wenig Bezug zur Waldnutzung. Rein räumliche Abfragen um Punkte sind somit nicht geeignet, um aussagekräftige Resultate zur Waldnutzung und -wahrnehmung zu erzielen. Wir stellen fest, dass es durch das LFI vorgegebene Punktstichprobenkonzept zu wenig georeferenzierte Daten auf den häufig genutzten sozialen Medienplattformen gibt, um detaillierte Analysen durchzuführen, und dies selbst in stark frequentierten Wäldern wie dem Uetliberg. Die LFI-Stichprobenpunkte stellen ein räumliches Stichprobenkonzept dar und sind nicht dort, wo die hohen Besucherströme zu erwarten sind, weswegen sich auch nicht alle LFI-Standorte für Befragungen eignen.

Erkenntnisse:

- ➔ Im Umkreis um die 10 zufällig gewählten Punkte wurden über 200'000 georeferenzierte Fotos aus Flickr und Instagram, sowie georeferenzierte Tweets gefunden
- ➔ Die unterschiedlichen Datenquellen sind stark miteinander korreliert
- ➔ Viele Daten aus sozialen Medien finden sich im weiteren Umkreis um die Stichprobenpunkte aber nur sehr wenige bis keine Daten im unmittelbaren Umkreis
- ➔ Für die Analyse von Waldnutzung und -erholung anhand sozialer Medien ist das Punktstichprobenkonzept aus dem LFI ungeeignet, da zu wenige Daten um die ausgewählten Punkte innerhalb von Wäldern vorkommen.

Flächige Analyse von Daten aus sozialen Medien in Schweizer Wäldern

Eine erste räumliche Analyse von Daten aus sozialen Medien rund um LFI-Stichprobenpunkte hat gezeigt, dass dieses Verfahren nicht zielführend ist. Anstelle von Stichproben um Punkte verwenden wir in einem zweiten Schritt deshalb die räumliche Ausdehnung von Schweizer Wäldern um zwei Fragen zu untersuchen:

- 1) Wie verhält sich die Verteilung von Daten aus sozialen Medien als Indikator für die tatsächliche Besuchsfrequenz zum Erholungspotential, welches anhand eines theoretischen Modells aus Wohnbevölkerung und Erreichbarkeit (siehe Brändli and Ulmer 2001) geschätzt wurde?
- 2) Wie lassen sich Erholungswälder anhand der von Nutzerinnen und Nutzern verwendeten Begriffe beschreiben?

Anhand des LFI-Datensatzes, der Waldflächen als Polygone abbildet, suchen wir Daten aus sozialen Medien innerhalb der LFI Polygone⁵. Für diese Analyse verwenden wir Daten von Twitter und Flickr, da nur diese Abfragen für ganze Flächen zulassen. Instagram lässt keine Abfragen nach Flächen zu, sondern liefert nur Daten im Umkreis von 5 Kilometern um gewählte Punkte, was zudem kostenpflichtig ist. Eine schweizweite Abfrage von Instagramdaten über unzählige Stichprobenpunkte (um so eine flächige Abdeckung zu erreichen) wäre möglich, aber würde den Kostenrahmen der meisten Forschungsprojekte sprengen.

Soziale Medien im Vergleich zum Erholungsnutzungspotential

Wir vergleichen nun die Abschätzung der Erholungsnutzung aufgrund sozialer Medien mit einem bereits bestehenden Modell, das aus dem LFI entwickelt wurde (Brändli and Ulmer 2001). Wir verwenden hierzu die Datensätze Twitter und Flickr wegen der Abdeckung der Gesamtfläche der Schweiz, die über Instagram nicht kostengünstig zu erreichen ist.

Bei Twitter finden wir schweizweit mehr Daten als bei Flickr. Dies deckt sich mit der vorherigen Analyse zur Datenverfügbarkeit um ausgewählte Stichprobenpunkte, wo wir mehr Daten auf Twitter als Flickr fanden. In Abbildung 10 ist die Konzentration der Verteilung um urbane Zentren in beiden Datensätzen ersichtlich, sowie generell die Ähnlichkeit der räumlichen Verteilung von Flickr und Twitter-Daten. Zudem enthalten die meisten Zellen des Erholungspotentialmodells keine Daten aus Flickr oder Twitter (in gelb dargestellt).

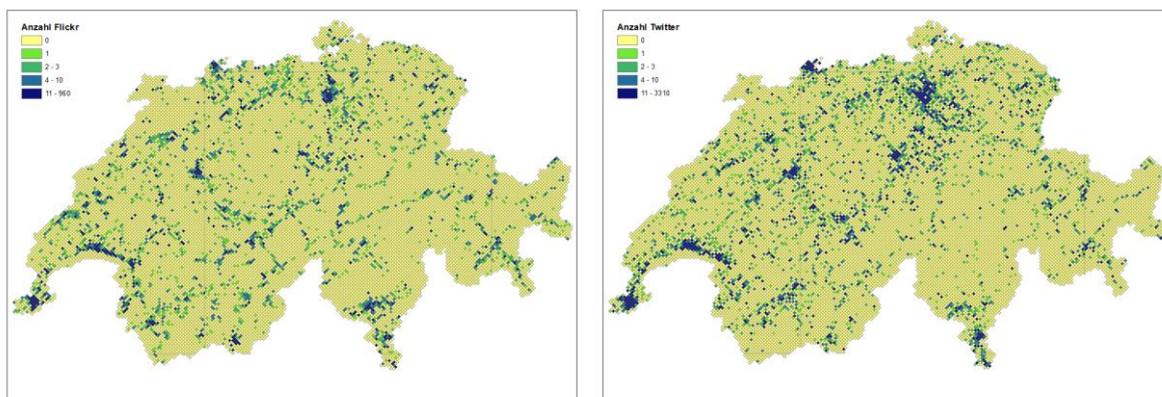


Abb. 10 Georeferenzierte Bilder auf Flickr (links) und Twitter (rechts)

Eine solche Werteverteilung in den Daten ist ungünstig für statistische Analysen, da ein Modell so aufgestellt wird, dass ein möglichst grosser Anteil der Daten erklärt werden kann. In unserem Beispiel weisen die meisten Standorte den Wert 0 (d.h. keine Daten) auf. Das Modell würde dann vor allem bei Zellen ohne Daten aus sozialen Medien gute Modellvorsagen liefern. Dies ist aber für die Beantwortung unserer Fragestellung nicht sinnvoll, weshalb wir die Daten in ein anderes Mass der Datenverfügbarkeit umrechnen. Wir berechnen die durchschnittliche Nähe von jedem Erholungspotentialstandort zu den jeweils 10 nächstgelegenen Flickr- bzw. Twitter-

⁵ Da der LFI-Datensatz sehr detailliert ist, und selbst kleinstflächige Waldstücke im Quadratmeterbereich abbildet, brachte dies einen enormen Rechenaufwand mit sich, um die Waldflächen mit den Millionen von Daten aus sozialen Medien für die ganze Schweiz zu verschneiden. Für zukünftige Forschung empfehlen wir die Verwendung von vereinfachten Waldpolygonen, um den massiven Rechenaufwand räumlicher Abfragen mit für die Fragestellung weitaus zu detaillierten Polygondatensätzen zu verringern.

Datenpunkten. Dies ergibt eine Werteverteilung von durchschnittlichen Distanzen zu sozialen Mediendaten für jede Rasterzelle in der Schweiz (Abb. 11). Diese Verteilung eignet sich nun besser zur Modellierung.

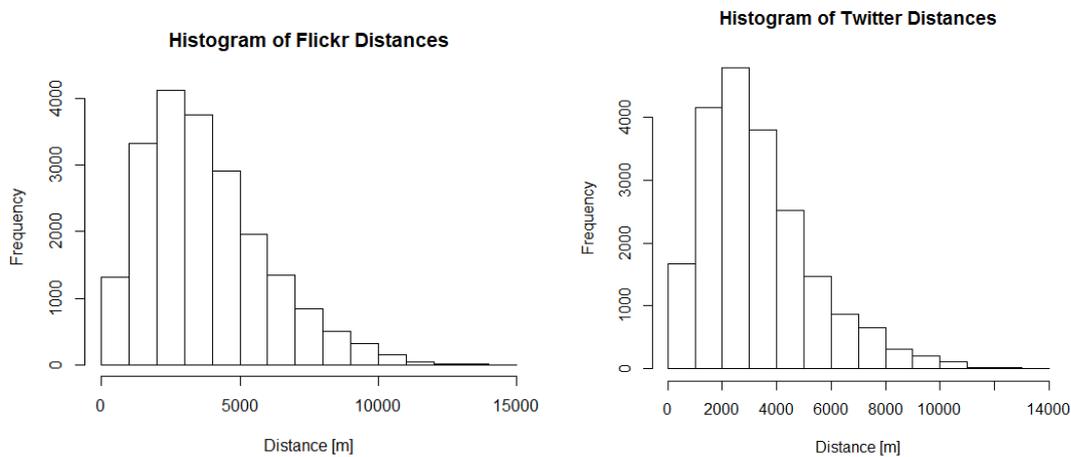


Abb. 11 Anzahl Rasterzellen des Erholungspotentialmodells mit Flickr (links) und Twitter-Daten (rechts)

Um den Zusammenhang zwischen Daten aus sozialen Medien und dem Erholungsnutzungspotential abzuschätzen, verwenden wir ein «Generalised Least Squared»-Regressionsmodell (Beguería and Pueyo 2009). Wir fanden einen signifikanten negativen Zusammenhang zwischen der durchschnittlichen Distanz zu den nächsten 10 Flickrdatenpunkten (Regressionskoeffizient = -0.865; p-value <0.001) als auch zu Twitterdaten (Regressionskoeffizient = -0.853; p-value < 0.001). Anders ausgedrückt zeigt das Modell: je näher die nächsten 10 Datenpunkte aus sozialen Medien sind, desto grösser ist das geschätzte Erholungsnutzungspotential.

Interessant sind in diesem Zusammenhang die Residuen, welche die Differenz zwischen der Schätzung aus sozialen Medien und dem Erholungsmodell angeben. Die Residuen treten in unserem Fall nicht zufällig im Raum auf, sondern sind stark räumlich korreliert (Abb. 12 u. 13).

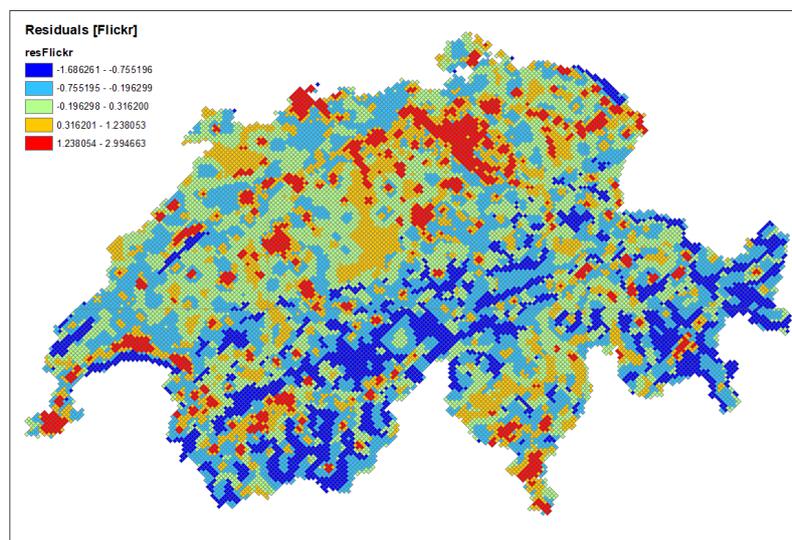


Abb. 12 Räumliche Verteilung der Residuen für Flickr

In Abb. 12 sind die um null verteilten Residuen grünlich eingefärbt. Dies sind Regionen, wo die Schätzungen aus den beiden Modellen gut übereinstimmen. Die positiven Residuen sind orange und rötlich eingefärbt. Dies sind Zellen, in denen wir mittels Flickrdaten die Nutzung tiefer einschätzen als das Erholungsnutzungsmodell. Dies sind vor allem Städte und angrenzende Gebiete wie Zürich, Basel, Genf, Lausanne und Locarno, wo die relative, aus Flickrdaten geschätzte Nutzung tiefer ist als die anhand der gesamten Wohnbevölkerung geschätzte. Im Gegensatz dazu zeigen die bläulich eingefärbten negativen Residuen Gebiete, wo wir mittels Flickrdaten die Erholungsnutzung höher einschätzen als mit dem Modell des Erholungsnutzungspotentials. Dies sind vor allem Gebiete in den Alpen, wo die Bevölkerungsdichte sehr tief ist, aber in umliegenden Zellen trotzdem Flickrfotos der touristisch attraktiven Alpen gemacht werden. Auffällig ist auch der Raum Flims, wo das Erholungsmodell die Nutzung im Vergleich zu den Flickrdaten erheblich höher einschätzt. Da das Erholungspotentialmodell auf der Wohnbevölkerung (inklusive Anzahl vermutlich oft leerstehender Ferienwohnungen) besteht, schätzt dieses Modell die tatsächliche Erholungsnutzung aber höher ein als das auf Flickrdaten mit tatsächlichen Beobachtungen basierende Modell.

Die räumliche Verteilung der Residuen mit Twitterdaten sieht sehr ähnlich aus wie bei Flickrdaten, was wiederum darauf hindeutet, wie ähnlich sich die Verteilung der Daten in diesen beiden Plattformen sind (Abb. 13).

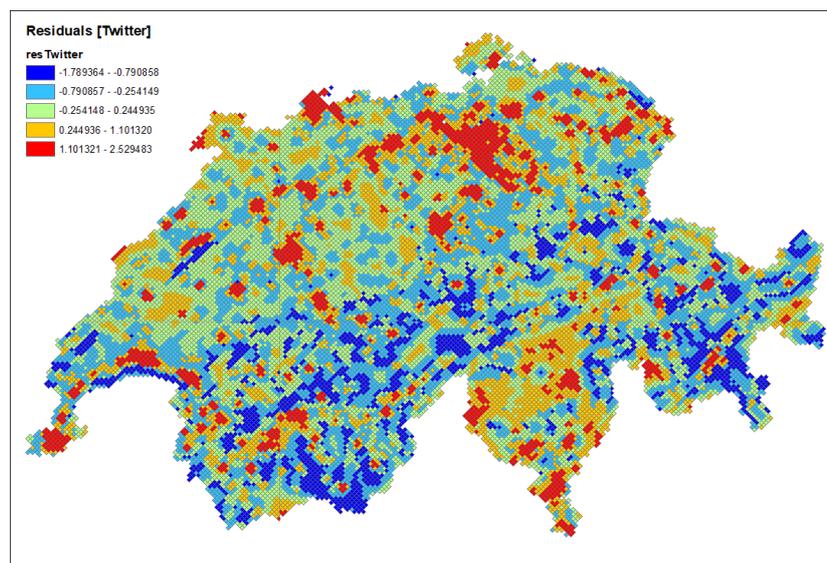


Abb. 13 Räumliche Verteilung der Residuen für Twitter

Wenn wir die Residuen noch mit der LFI-Waldmaske verschneiden, sehen wir, in welchen Waldgebieten wir anhand sozialer Medien die Nutzung höher oder tiefer einschätzen als im Erholungspotentialmodell, hier am Beispiel Flickr (Abb. 14):

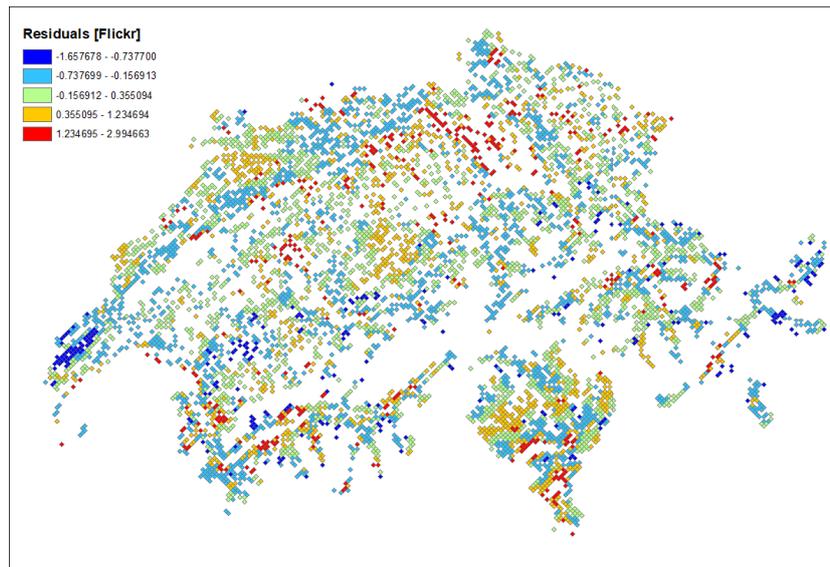


Abb. 14 Räumliche Verteilung der Residuen für Flickr innerhalb von LFI Waldpolygonen

Da die Residuenkarten eine starke räumliche Autokorrelation⁶ aufzeigen, wird unser Modell die Korrelation überschätzen. Für die Waldgebiete berechnen wir erneut ein Modell, welches explizit die räumliche Autokorrelation der Daten berücksichtigt⁷. Der Zusammenhang zwischen Modell aus sozialen Medien und Erholungspotentialmodell bleibt signifikant negativ (Flickr: -0.651; Twitter: -0.6881) und etwas tiefer, als diejenigen Werte, die wir ohne Korrektur für räumliche Autokorrelation berechnen. Die neu berechneten Korrelationswerte sind sogar leicht höher für Standorte innerhalb von LFI-Polygonen (Flickr: -0.7011; Twitter: -0.7129) als wenn alle Standorte auch ausserhalb des Waldes miteingerechnet werden (Flickr: -0.651; Twitter: -0.6881). Dies deutet darauf hin, dass innerhalb von Waldpolygonen die Daten aus sozialen Medien das Erholungspotential etwas besser schätzen als wenn der gesamte Datensatz in Betracht gezogen wird. Da die LFI Waldflächen keinen Siedlungsraum enthalten, wo wir grosse Unterschiede zwischen den Modellen festgestellt hatten (Abb. 13) wird die Gesamtkorrelation leicht besser, wenn diese Flächen nicht mitgerechnet werden.

Die Resultate zeigen, dass wir trotz unterschiedlicher Datenverfügbarkeit zwischen Flickr und Twitter mit beiden Datensätzen «Hot- und Coldspots» von Erholungsnutzung gesamtschweizerisch abschätzen können, was ein Mehrwert im Vergleich zur bisherigen theoretischen Abschätzung bedeutet.

Allerdings weisen wir darauf hin, dass das Modell des Erholungspotentials stark auf die Wohnbevölkerungsdichte fokussiert (Brändli and Ulmer 2001). Flickr und Twitter-Daten korrelieren ebenfalls sehr stark mit der Bevölkerungsverteilung (urbane Zentren sind in sozialen Medien viel stärker abgebildet als rurale Gebiete). Die Übereinstimmungen im Modell sind daher möglicherweise auf die Übereinstimmung zwischen Bevölkerungsdichte und Dichte an tweets und Flickr-Bildern zurückzuführen.

⁶ Räumliche Autokorrelation bedeutet, dass nahe Punkte mit hoher Wahrscheinlichkeit ähnlichere Werte aufweisen als weit entfernte. Das Vorhandensein räumlicher Autokorrelation verletzt die Grundannahme des Modells, und muss deshalb korrigiert werden.

⁷ Dieses Modell (Rational Quadratic) berücksichtigt explizit die räumliche Autokorrelation. Da der Rechenaufwand sehr hoch ist, führten wir zur Bestimmung des Zusammenhangs jeweils 10 Iterationen mit je 20% aller Daten durch, um so den Rechenaufwand zu minimieren und geben hier den durchschnittlich ermittelten Regressionskoeffizienten an.

Wichtig zu erwähnen ist an dieser Stelle, dass weder das eine noch das andere Modell einen tatsächlichen Wert darstellt. Beide Modelle geben lediglich eine Schätzung. Durch eine Kombination der beiden Modelle liesse sich aber möglicherweise ein robusteres Modell entwickeln, welches nicht nur das Potential einschliesst (GIS-Modell), sondern auch die tatsächlich beobachtete Nutzung (anhand sozialer Medien).

Erkenntnisse:

- ➔ Besuchsfrequenzen lassen sich über lange Zeiträume und grosse Untersuchungsgebiete anhand von Daten aus sozialen Medien abschätzen.
- ➔ Das Erholungspotentialmodell aus dem LFI schätzt die Erholungsnutzung in urbanen Gebieten höher ein als Modelle mit Flickr oder Twitter-Daten.
- ➔ In alpinen Räumen mit geringer Bevölkerungsdichte aber hohem Tourismusaufkommen schätzen Modelle mit Flickr oder Twitter-Daten die Besuchsfrequenz höher ein als das Erholungspotentialmodell.

Wälder anhand von Beschreibungen aus sozialen Medien charakterisieren

In einem letzten Schritt verwenden wir die Wörter, die in den Bildbeschreibungen auf sozialen Medienplattformen wie Instagram und Flickr vorhanden sind, um Rückschlüsse auf die Wahrnehmung von Wäldern zu ziehen. NutzerInnen verwenden Bildbeschreibungen (hashtags oder tags), damit ihre Bilder von anderen gefunden werden können. Diese Beschreibungen übernehmen daher eine wichtige Funktion und werden von den meisten Nutzerinnen und Nutzern gewissenhaft vorgenommen (Hollenstein and Purves 2010). Anhand solcher tags lassen sich deshalb Orte, oder in unserem Fall Wälder, aus nutzergenerierten Textdaten beschreiben. Um Erholungswälder mit tags zu charakterisieren, wählen wir exemplarisch drei Waldgebiete aus: Zürich (Uetliberg), Neuchâtel und Locarno.

Da wir auf Instagram die meisten Daten um diese Standorte finden, verwenden wir für diese Analyse beispielhaft diese Plattform. Analysen von Textinhalten sind aber auch mit Twitter und Flickr, sowie mit weiteren Quellen wie Wanderblogs etc. möglich. Von jedem Standort analysieren wir zuerst alle Instagram-Daten im Umkreis von fünf Kilometern. Als nächstes wählen wir nur die Daten aus, die innerhalb dieses Umkreises auch in LFI Waldpolygonen liegen. Die so extrahierten tag-Daten liegen als Wortlisten vor, aus denen wir sogenannte «Stopwörter» entfernen. Stopwörter sind Wörter, die in der Sprache sehr häufig vorkommen, die aber für unsere Forschungsfrage keinen Erkenntnisgewinn bringen. Wir filtern sie deshalb heraus. Dazugehörige Wortgruppen sind zum Beispiel Pronomen wie «ich, sie, er, wir», Artikel («der, die das») oder Konjunktionen («und»). Diese Art der Filterung ist ein gängiges Verfahren im Gebiet des «Natural Language Processing» (Manning and Schütze 1999). Anschliessend zählen wir die Anzahl Nennungen der einzelnen Wörter. Diese Listen mit Häufigkeiten können wir anhand Wortwolken visualisieren.



Abb. 15 Wortwolken von Instagram Bildbeschreibungen aus Locarno
(links: ganzer Umkreis; rechts: innerhalb von LFI Waldpolygonen)

Obwohl das so dargestellte Mass der Häufigkeit von Begriffen einfach verständlich ist, werden oft komplexere Methoden empfohlen, um die Wichtigkeit von Wörtern zu eruieren. So zum Beispiel das *tfidf*⁸-Mass aus dem Fachbereich *Information Retrieval*, welches verwendet wird, um zum Beispiel Internetseiten nach ihrer Wichtigkeit zu ordnen. Das *tfidf*-Mass findet auch Anwendung in der Analyse von Daten aus sozialen Medien für die Charakterisierung von Orten und Landschaften (Purves and Mackaness 2016; Derungs and Purves 2013). Die Berechnung des *tfidf*-Masses übersteigt allerdings den Umfang dieser Pilotstudie. Daher zeigen wir am relativ einfach verständlichen Beispiel der Häufigkeit, wie wir anhand von Begriffen Wälder anhand von tags charakterisieren können.

Bei der Analyse von tags aus sozialen Medien fällt auf, wie prominent Ortsnamen in den Bildbeschreibungen vorkommen (Abb. 15). Diese Tatsache wurde bereits ausgiebig erforscht und dient unter anderem dazu, unscharfe Räume wie Stadtkerne oder Landschaften mittels georeferenzierter Bilder einzugrenzen (Wartmann, Acheson, and Purves 2018; Hollenstein and Purves 2010; Jones et al. 2008). In dieser Studie sind wir allerdings nicht daran interessiert, bestimmte Wälder anhand von Ortsnamen abzugrenzen. Daher sind Ortsnamen für die vorliegende Analyse nicht relevant. In unseren Daten waren aber wie am Beispiel Locarno ersichtlich Ortsnamen sowohl im gesamten Umkreis (Abb. 15 links) als auch nur innerhalb des Waldes (Abb. 15 rechts) dominant. Da die Häufigkeiten von Ortsnamen derart hoch ist, verschwinden im Gegensatz dazu für uns interessante Begriffe derart, dass sie fast nicht erkennbar sind, wie zum Beispiel *nature*, *amazing*, *panorama* (Abb 15. mitte). Da die Listen zudem viele Begriffe enthalten, die wir nicht für die Beschreibung von Landschaften verwenden wollen (z.B. *instagood*, *picoftheday*, *myswitzerland*, *ticinoturismo*) filtern wir die Begriffe noch zusätzlich. Dazu gibt es grundsätzlich zwei Möglichkeiten.

⁸ *Term-frequency-inverse-document-frequency* ist ein Mass zur Beurteilung der Relevanz von Termen in Dokumenten einer Kollektion aus verschiedenen Dokumenten

Uetliberg		Neuchâtel		Locarno	
Begriff	Häufigkeit	Begriff	Häufigkeit	Begriff	Häufigkeit
winter	11	beautiful	4	autumn	19
snow	9	nature	3	lake	15
cars	8	lake	3	view	12
christmas	7	tunnel	3	snow	11
autumn	7	forest	2	sun	11
love	6	trees	2	beautiful	10
travel	6	light	2	day	8
trip	5	autumn	2	nature	8
walking	5	walk	2	happy	7
nature	5			trees	7
sunset	5			winter	6
sport	5			landscape	6
beautiful	4			love	5
day	4			sunshine	4
landscape	4			mountain	4
beauty	3			sky	4
cycling	3			dog	4
girl	3			mountains	4
morning	3			sunrise	4
forest	3			weekend	4

Tab. 1 Häufigste zwanzig landschaftsrelevante Begriffe für die Wälder im Umkreis vom Uetliberg, Neuchâtel und Locarno

Solche Listen aus Wäldern lassen sich mit Listen aus dem gesamten Datensatz vergleichen. So liessen sich Unterschiede in der Erholungsnutzung innerhalb von Wäldern im Vergleich zu ausserhalb von Wäldern feststellen, indem wir die erwähnten Wörter und ihre Häufigkeiten für die Kategorie «Aktivitäten» vergleichen. Solche Listen lassen sich weiters mit Methoden aus der maschinellen Sprachverarbeitung auch analysieren, um Ähnlichkeiten zwischen Wäldern aufgrund ihrer Beschreibungen zu finden. Dieses Verfahren wurde bereits erfolgreich auf Beschreibungen in sozialen Medien von Schweizer Erholungslandschaften angewandt, Berg-, Fluss, Seen-, Moor- und Hügellandschaften zu differenzieren (Wartmann, Acheson, and Purves 2018).

Die vorliegende erste Analyse von Inhalten aus sozialen Mediendaten in Wälder zeigt, dass wir anhand von Visualisierungen mit Wortwolken und Begriffslisten erste Einblicke in Aktivitäten von Waldnutzerinnen und -nutzern gewinnen können. Dies auch in Wäldern und Zeiträumen, bei denen wir über keine lokalen Beobachtungen verfügen.

Erkenntnisse:

- ➔ Textinhalte aus sozialen Medien eignen sich dazu, Wälder aufgrund ihrer von Nutzerinnen und Nutzern wahrgenommenen Eigenschaften zu beschreiben.
- ➔ Solche Beschreibungen können für Vergleiche zwischen der Waldnutzung in verschiedenen Gebieten herbeigezogen werden.

Evaluation von Daten aus sozialen Medien für die Erforschung von Waldnutzung und -wahrnehmung und praktische Empfehlungen

Aufgrund der Literaturrecherche und unserer empirischen Pilotstudie gehen wir in diesem Kapitel kritisch auf das Potential solcher Daten zur Erforschung der Erholungsnutzung in Schweizer Wäldern ein und geben Empfehlungen ab für die Nutzung solcher Daten. Dabei gehen wir in jedem Abschnitt auf eine Frage ein, welche wir basierend auf der Literaturrecherche und der empirischen Pilotstudie beantworten.

Wie gut ist die Verfügbarkeit und die Abdeckung von Daten aus sozialen Medien?

Obwohl gewisse Studien das Potential sozialer Medien für die Wissenschaft herausstreichen und beispielsweise die Behauptung aufgestellt haben, dass anhand sozialer Medien Erholungsverhalten auf der ganzen Welt analysiert werden kann (Wood et al. 2013), ist diese Aussage kritisch zu betrachten. Neuere, detailliertere Studien zeigen, dass die Datenverfügbarkeit in sozialen Medien sehr ungleich ist: eine hohe Konzentration an Daten finden wir im urbanen Raum und an hochfrequentierten Orten entlang von Strassen oder in bekannten Nationalparks und Freizeitattraktionen (Dunkel 2015). Je nach Studiengebiet und Themenfokus ist die Datenverfügbarkeit in den sozialen Medien sehr spärlich und wenig besuchten Orten finden wir teilweise sehr wenig bis gar keine Daten (Levin, Lechner, and Brown 2017). Unsere Schweizer Pilotstudie bestätigt dies. Obwohl wir für die ganze Schweiz sehr viele Daten finden, verringert sich die Verfügbarkeit bei der Betrachtung spezifischer Orte und nur innerhalb von Wäldern massiv. Eine geringe Datenverfügbarkeit kann dazu führen, dass die Korrelation von beobachteten Besuchszahlen und aus den sozialen Medien ermittelten Werten geringer ist, und sich dadurch die Nutzbarkeit solcher Daten für das Landschaftsmanagement reduziert (Heikinheimo et al. 2017). Bei Betrachtung kleinerer Studiengebiete und mit einem Fokus auf ein spezifisches Thema können wir daher nicht von «big data» sprechen, sondern eher von «small data». In der Tat verwenden gewissen Studien zu sozialen Medien Stichprobengrößen, die gut mit einer Umfrage hätten erreicht werden können (z.B. Guerrero et al. 2016).

Empfehlung. Wir raten von der ausschliesslichen Nutzung von sozialen Medien für stark räumlich oder zeitlich eingegrenzte Studien ab. Das Potential von Daten aus sozialen Medien sehen wir in deren Verwendung für Quervergleiche oder als Ergänzung zu traditionellen Erhebungsmethoden wie Umfragen oder *in situ* Zählungen. Zudem bietet die Analyse von sozialen Medien auch das Potential, hochfrequentierte und daher wertgeschätzte Orte und Landschaften zu identifizieren, um beispielsweise ein Stichprobendesign zu entwerfen, welches solche Orte miteinbezieht. Allerdings muss beachtet werden, dass Orte, an denen viele Fotos hochgeladen werden (z.B. Ausblick, spezielles Fotosujet), möglicherweise fotogener sind als andere Orte (z.B. Waldweg, dichte Vegetation), die deswegen nicht weniger stark frequentiert werden, aber weniger fotografiert. Zudem bilden soziale Medien vermutlich eher eine touristische Erholungsnutzung ab, da Alltagserholung wie mit dem Hund im Wald spazieren zu gehen eher nicht geteilt wird. Je nach Plattform ist es aber durchaus möglich, generell auch

Alltagserholung zu erfassen, wie beispielsweise Jogging im Wald über Fitness-Tracking Apps wie Strava.

Für zukünftige Forschung empfehlen wir auf Untersuchungen auf grober Skala zu fokussieren (z.B. grosse Untersuchungsräume oder sehr populäre Orte) und für fein skalierte Untersuchungen (wie z.B. Anzahl NutzerInnen auf spezifischen Wegen) herkömmliche Monitoringinstrumente zu verwenden. Messungen von solchen Instrumenten wie Schrittzähler können zudem an ausgewählten stark frequentierten Standorten direkt mit Daten aus sozialen Medien abgeglichen werden. Dies wird zeigen, ob soziale Medien allenfalls als Ersatz bei stark frequentierten Orten eingesetzt werden können. Für niedrig frequentierte Orte oder Wälder empfehlen wir weiterhin die Nutzung von herkömmlichen Monitoringinstrumenten, da bei tiefer Frequenz der Einfluss von einzelnen NutzerInnen oder Nutzern sehr hoch ist und die Ergebnisse stark beeinflusst. Eine weitere Möglichkeit ist die Kombination von "klassischen" Instrumenten mit Daten aus sozialen Medien, um dadurch robuste Modelle zu generieren, die dann auch auf dünn besiedelte Gebiete angewendet werden könnten.

Wie repräsentativ sind Daten aus sozialen Medien?

Immer mehr Menschen verwenden soziale Medien in ihrem Alltag und dokumentieren damit ihre Freizeitaktivitäten, woraus sich Erkenntnisse in Bezug auf die Erholungsnutzung ziehen lassen. Allerdings stellen sich Fragen der Repräsentativität solcher Daten (Tufekci 2014). Generell wird die Annahme getroffen, dass eher jüngere, gut ausgebildete mit hohem Einkommen soziale Medien nutzen (Li, Goodchild, and Xu 2013). Empirische Untersuchungen zeichnen je nach Kontext ein differenzierteres Bild. Eine Umfrage im südafrikanischen Krüger-Nationalpark ergab zum Beispiel, dass über 80% der befragten BesucherInnen mindestens eine Plattform nutzen (Hausmann et al. 2018). Allerdings fanden sich je nach Plattform soziodemographische Unterschiede. So waren Twitter, Instagram und Facebook eher bei Jüngeren Personen beliebt, während sich im Krügerpark bei der Verwendung von Flickr keine Altersunterschiede zeigten. Zudem war je tiefer das angegebene Einkommen der Befragten, desto höher die Nutzung sozialer Medien (Hausmann et al. 2018), was die Annahme widerlegt, dass soziale Mediennutzung und ein hohes Einkommen zusammenhängen (Li, Goodchild, and Xu 2013). In einem finnischen Nationalpark zeigte sich ausserdem, dass nur wenige BesucherInnen, welche Fragebögen ausfüllten, überhaupt soziale Medien nutzen, obwohl für den Park viele Daten verfügbar sind (Heikinheimo et al. 2017). Dies deutet darauf hin, dass NutzerInnen sozialer Medien nicht an den Befragungen teilnahmen, während Teilnehmende der Befragungen keine sozialen Medien verwenden. Wir decken also unterschiedliche Bevölkerungsgruppen mit unterschiedlichen Methoden ab, was deren Kombination nahelegt. Einzelne Personen können einen grossen Einfluss auf Daten haben, weil sie überproportional aktiv auf sozialen Medien sind und viel Inhalte liefern. Eine solche Beeinflussung lässt sich aber gut über Filtermethoden beheben, welche die Anzahl Bilder pro Nutzer einschränken. Zudem ist es einfach, identische taglisten zu filtern, welche auf sogenannte bulk-uploads hindeuten, wenn jemand hunderte Bilder mit identischen Informationen (tags und Koordinaten) hochlädt.

Empfehlung. Durch eine Kombination von Methoden der Analyse sozialer Medien und direkter Befragungen lässt sich ein besseres Gesamtbild der Diversität an Erholungsnutzenden gewinnen. Je nach Plattform ermöglichen es soziale Medien, an NutzerInnen und Nutzer zu

gelangen, die nicht an traditionellen Befragungen teilnehmen und in solchen Erhebungsinstrumenten unterrepräsentiert sind. Mit Befragungen erreicht man hingegen gut Nutzergruppen, die keine sozialen Medien verwenden. Durch Filterung können wir eine starke Überrepräsentation einzelner Nutzerinnen und Nutzer vermeiden.

Wie glaubwürdig sind Daten aus sozialen Medien?

Ein grosses Augenmerk wurde auf die Glaubwürdigkeit der Daten aus sozialen Medien gelegt (Flanagin and Metzger 2008; Spielman 2014). Hier sind Vergleiche zwischen OpenStreetMap Karten mit offiziellen Karten zu erwähnen, die zeigen, dass Nutzerinnen und Nutzer relativ genaue Daten liefern (Haklay 2010; Haklay et al. 2010). Für Flickrdaten wurde eine Überprüfung der Standortgenauigkeit und der verwendeten Tags vorgenommen (Hollenstein and Purves 2010). Dabei zeigte sich, dass die NutzerInnen eine hohe Standortgenauigkeit erreichen und «sinnvolle» tags verwenden, obwohl auch unbeabsichtigte Fehler auftreten können (z.B. wird der Käferberg in Zürich als der bekanntere Üetliberg getagged). Da die tags der Auffindbarkeit der Daten dienen, die wiederum nötig ist, um «likes» und Kommentare zu erhalten, finden wir relativ wenig nicht-sinnvolle oder semantisch falsche tags. Beispielsweise zeigt eine Suche nach dem tag «Wald» auf Flickr tatsächlich Waldfotos und nicht beispielsweise Partyfotos, da dies den NutzerInnen keinen Vorteil bringen würde. Solange die Nutzung dieser Daten für die Forschung nicht sehr bekannt ist, gehen wir auch nicht von einer aktiven Beeinflussung der Daten aus.

Empfehlung. Wir gehen basierend auf der Literatur und durch die erwünschte Auffindbarkeit des Materials auf sozialen Medien davon aus, dass Nutzerinnen und Nutzer mehrheitlich gewissenhaft taggen und ihre Daten mit Koordinaten versehen, was die Daten relativ glaubwürdig macht. Wir empfehlen in jedem Fall die Daten stets kritisch zu überprüfen und auch einen Teil des Materials komplett zu sichten (z.B. Bilder und deren tags in einem ausgewählten Standort betrachten), da auch unerwartete Resultate auftreten können, wie «tweets aus Justin Biebers Herz» (Hecht et al. 2011).

Wer hat die Kontrolle über den Prozess und die Verfügbarkeit der Daten?

Da Nutzerinnen und Nutzer die Daten generieren, unterliegt der Prozess der Informationsproduktion keinerlei Kontrolle durch die Wissenschaft. Dies ist einerseits ein Potential dieser Daten, da wir eine direkte Beeinflussung durch Forschende ausschliessen können, und die Daten so eher ein tatsächliches Verhalten wiedergeben (obwohl soziale Erwünschbarkeit gegenüber anderen NutzerInnen auch hier von Belang ist). Die Fotos werden von NutzerInnen während ihrer Aktivitäten gemacht. Die Fotografierenden sind sich der Nutzung ihrer Daten für die Erforschung von Erholung sehr wahrscheinlich nicht bewusst. Die Daten sind dadurch eine bessere Annäherung an die unbewusste Landschaftserfahrung (Dunkel 2015). Andererseits schafft diese unbeeinflusste Produktion von Daten grosse Herausforderungen. Die Popularität von Plattformen wandelt ständig, ohne dass Forscherinnen oder Forscher darauf einen Einfluss haben. Gewisse Plattformen haben eine kurze Lebensdauer, und zeichnen sich durch einen eigentlichen Hype gefolgt von starker Abflachung in der Nutzung aus, weshalb sie sich nicht für langfristiges Monitoring eignen (Norman and Pickering 2017). Selbst bei grossen Firmen wie Google ist die Datenverfügbarkeit über längere

Zeit nicht gegeben, wie das Beispiel von Panoramio zeigt, dessen Service Google im 2016 eingestellt hat¹⁰. Ein weiteres Problem dieser Plattformen ist, dass die Dateninhaberschaft meist bei kommerziellen Firmen liegt, welche die Nutzungsbedingungen der Daten praktisch über Nacht ändern können, und die Daten nicht mehr kostenlos zur Verfügung stehen. Bei Twitter und Instagram sind nur gewisse Abfragen kostenfrei, und selbst bei diesen stark eingeschränkten Abfragemöglichkeiten gibt Twitter nur rund 1% der verfügbaren Daten gratis ab, der Rest muss teuer erworben werden. Die Businessmodelle dieser Firmen können auch ändern. So wurde die bisher frei verfügbare Plattform Flickr erst kürzlich von «SmugMug» aufgekauft. Es bleibt abzuwarten, ob sich dadurch das Businessmodell von Flickr ändert, und die Daten kostenpflichtig oder allenfalls ganz von der Nutzung für die Forschung ausgeschlossen werden. Ein weiteres Problem ist, dass sich die Akzeptanz in der Bevölkerung über die Zeit ändern kann, was sowohl die Produktion/Erfassung von Daten als auch deren Nutzung beeinflusst. Zudem lässt sich momentan noch kaum abschätzen, wie Veränderung der Rahmenbedingungen (z.B. der General Data Protection Act GDPR) sich auswirken werden.

Empfehlung. Forschung, die sich mit sozialen Medien beschäftigt, muss im Bewusstsein geschehen, dass die Datenquellen von einem auf den anderen Tag versiegen können. Wir empfehlen daher längerfristig, selbst in der Erfassung von Daten aktiv zu werden. Wo selbst Daten aus sozialen Medien generiert werden, gibt es keine Abhängigkeit von kommerziellen Plattformen mehr. Ein grosses Potential sehen wir bei unabhängigen Plattformen, die von Bürgerinnen und Bürgern (mit oder ohne direkte Zusammenarbeit mit Forschenden) im Sinne der Citizen Science erzeugt werden. Ein gutes Beispiel hierfür ist die Plattform «Geograph» aus Grossbritannien, welche von Freiwilligen erstellt und unterhalten wird, und sämtliche Daten wie georeferenzierte Landschaftsbilder und -beschreibungen unter einer Creative Commons Lizenz zur Verfügung stellt. Während Geograph von freiwilligen Lehrkräften ohne Mitarbeit von Forschenden erstellt wurde, empfehlen wir, für die Forschung und das Monitoring vermehrt in die aktive Erfassung von Daten durch soziale Medien zu investieren, welche von Forschenden begleitet wird. Dies kann einerseits mit eigens entwickelten Plattformen sein, z.B. eine App, über die Nutzerinnen und Nutzer Waldaktivitäten eingeben oder Fotos von ihrem Waldbesuch hochladen. Die Nutzenden könnten im Wald selbst über QR-Codes oder Informationstafeln mit Download-Links auf die App aufmerksam gemacht werden. Andererseits kann eine solche Erfassung auf bestehende Plattformen aufbauen, die bisher rein auf Informations*vermittlung* setzten, wo neu eine Komponente zur Informations*erfassung* eingebaut werden könnte. Schliesslich sind auch Ansätze möglich, die auf spielerische Art Daten erfassen. Die Masterarbeit von Manuel Bär am Geographischen Institut der UZH zeigt das Potential zur Erfassung von Landnutzung anhand eines Online-Spiels «StarBorn». In einem intergalaktischen Team erkunden darin SpielerInnen mit ihren Smartphones die Erde und klassieren Landnutzung. Sie gewinnen so «Sternenstaub», der ihre eigene sowie die Wertung des Teams verbessert. Solche «gamifizierten» Ansätze haben das Potential, schnell und gezielt Daten zu sammeln von Nutzergruppen, die sich nicht an eher forschungsorientierten Citizen-Science Ansätzen beteiligen würden. Ein solcher Ansatz erfordert eine gezielte Begleitung durch die Forschung, sowie eine aktive Anwerbung von Spielerinnen und Spielern über verschiedene Kanäle, um eine angemessene Anzahl an Teilnehmenden zu erreichen.

¹⁰ <https://www.panoramio.com>, zuletzt aufgerufen am 15.5.2018

Schlussbetrachtung

Die Literaturrecherche zeigt, dass soziale Medien bereits zur Analyse von Besuchsaufkommen, Besucherverhalten und Wahrnehmung von Naturräumen eingesetzt wird. Durch die testweise Analyse konnten wir für das Untersuchungsgebiet Schweiz zeigen, dass solche Analysen unter gewissen Bedingungen ebenfalls möglich ist.

Die empirische Analyse hat klar gezeigt, dass ein vorgegebenes Stichprobenraster nicht geeignet ist, um für zufällig ausgewählte Punkte eine angemessene Datengrundlage aus sozialen Medien zu extrahieren. Eine direkte Anbindung an das Waldmonitoring LFI über die einzelnen Stichprobenpunkte ist deshalb nicht möglich. Erfolgsversprechender sind Ansätze, welche die gesamte Waldfläche über längere Zeiträume betrachten, um dadurch die Nutzung abzuschätzen. Ein Vergleich des Erholungspotentialmodells aus dem LFI (Brändli and Ulmer 2001) mit einer aus sozialen Mediendaten errechneten Schätzung der Nutzung zeigt Unterschiede in den beiden Modellen auf. Das Erholungspotentialmodell schätzt dabei die Erholungsnutzung in urbanen Räumen aufgrund der Wohnbevölkerung höher ein als die Modelle mit sozialen Medien. In alpinen Räumen mit wenig Bevölkerung schätzen die Modelle anhand sozialer Medien hingegen die Nutzung höher ein. Eine Kombination beider Modellansätze scheint deshalb vielversprechend.

Ein grosses Potential liegt auch in der Analyse von Textinhalten von sozialen Medien, wodurch wir anhand von Sprache Wälder und unternommene Aktivitäten in Wäldern zwischen verschiedenen Gebieten vergleichen können. Der semantische Bestandteil in sozialen Mediendaten machen diese Informationen ausserdem interessant für die Beobachtung der Landschaftswahrnehmung, wie sie beispielsweise über Wahrnehmungsindikatoren im LABES-Monitoring geschieht. Für Indikatoren wie «Wahrgenommene Schönheit der Landschaft» oder «Faszination» bieten sich Möglichkeiten, diese mit Daten aus sozialen Medien zu vergleichen.

Da momentan eine grosse Abhängigkeit von kommerziellen Firmen besteht, welche Daten aus sozialen Medien anbieten, empfehlen wir für zukünftige Forschung und Anwendung in diesem Bereich auf eigens entwickelte Plattformen zu setzen, deren Beständigkeit über die Zeit garantiert werden kann, und worüber die Zugänglichkeit zu den Daten garantiert ist. Dies ist bei Anbietern wie Flickr, Instagram und Twitter nicht der Fall, wodurch die Nützlichkeit solcher Daten für langfristige Beobachtungsreihen verringert. Solche eigens entwickelten Plattformen können im Sinne von Citizen Science unter aktiver Mitwirkung der Bevölkerung erstellt werden, bei denen Personen für die Forschung interessante Daten liefern. Hier sind verschiedene Ansätze im Bezug auf die Waldnutzung vorstellbar und empirisch zu prüfen.

Literaturangaben

- Barry, Sheila J. 2014. "Using Social Media to Discover Public Values, Interests, and Perceptions about Cattle Grazing on Park Lands." *Environmental Management* 53 (2). Springer: 454–64.
- Beguería, S, and Y Pueyo. 2009. "A Comparison of Simultaneous Autoregressive and Generalized Least Squares Models for Dealing with Spatial Autocorrelation." *Global Ecology and Biogeography* 18 (3). Wiley Online Library: 273–79.
- Brändli, Urs-Beat, and Ulrich Ulmer. 2001. "Recreational Function." In *Swiss National Forest Inventory: Methods and Models of the Second Assessment*, edited by Peter Brassel and Heike Lischke. Birmensdorf: WSL.
- Campelo, Maria B., and Ricardo M. Nogueira Mendes. 2016. "Comparing Webshare Services to Assess Mountain Bike Use in Protected Areas." *Journal of Outdoor Recreation and Tourism* 15 (October): 82–88. doi:10.1016/j.jort.2016.08.001.
- Casalegno, Stefano, Richard Inger, Caitlin DeSilvey, and Kevin J Gaston. 2013. "Spatial Covariance between Aesthetic Value & Other Ecosystem Services." Edited by Tobias Preis. *PLoS ONE* 8 (6): e68437. doi:10.1371/journal.pone.0068437.
- Chua, Alvin, Loris Servillo, Ernesto Marcheggiani, and Andrew Vande Moere. 2016. "Mapping Cilento: Using Geotagged Social Media Data to Characterize Tourist Flows in Southern Italy." *Tourism Management* 57 (December): 295–310. doi:10.1016/j.tourman.2016.06.013.
- Daume, Stefan, Matthias Albert, and Klaus von Gadow. 2014. "Forest Monitoring and Social Media—Complementary Data Sources for Ecosystem Surveillance?" *Forest Ecology and Management* 316. Elsevier: 9–20.
- Derungs, Curdin, and Ross S Purves. 2013. "From Text to Landscape: Locating, Identifying and Mapping the Use of Landscape Features in a Swiss Alpine Corpus." *International Journal of Geographical Information Science* 28 (6): 1272–93.
- Donahue, Marie L, Bonnie L Keeler, Spencer A Wood, David M Fisher, Zoé A Hamstead, and Timon McPhearson. 2018. "Using Social Media to Understand Drivers of Urban Park Visitation in the Twin Cities, MN." *Landscape and Urban Planning* 175. Elsevier: 1–10.
- Dunkel, Alexander. 2015. "Visualizing the Perceived Environment Using Crowdsourced Photo Geodata." *Landscape and Urban Planning* 142 (October): 173–86.
- English, Donald B.K, Susan M. Kocis, Stanley J. Zarnoch, and Ross J. Arnold. 2002. "Forest Service National Visitor Use Monitoring Process: Research Method Documentation." *General Technical Report (GTR)-SRS-057*. Asheville, NC. <https://www.srs.fs.usda.gov/pubs/5000/>, accessed 7.5.2017.
- Figueroa-Alfaro, Richard Wagner, and Zhenghong Tang. 2017. "Evaluating the Aesthetic Value of Cultural Ecosystem Services by Mapping Geo-Tagged Photographs from Social Media Data on Panoramio and Flickr." *Journal of Environmental Planning and Management* 60 (2): 266–81. doi:10.1080/09640568.2016.1151772.
- Flanagin, Andrew J, and Miriam J Metzger. 2008. "The Credibility of Volunteered Geographic

- Information." *GeoJournal* 72 (3–4). Springer: 137–48.
- Guerrero, Paulina, Maja Steen Møller, Anton Stahl Olafsson, and Bernhard Snizek. 2016. "Revealing Cultural Ecosystem Services through Instagram Images: The Potential of Social Media Volunteered Geographic Information for Urban Green Infrastructure Planning and Governance." *Urban Planning* 1 (2): 1. doi:10.17645/up.v1i2.609.
- Haklay, Mordechai. 2010. "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets." *Environment and Planning B: Planning and Design* 37 (4). SAGE Publications Sage UK: London, England: 682–703.
- Haklay, Mordechai, Sofia Basiouka, Vyron Antoniou, and Amer Ather. 2010. "How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information." *The Cartographic Journal* 47 (4). Taylor & Francis: 315–22.
- Hausmann, Anna, Tuuli Toivonen, Rob Slotow, Henriikki Tenkanen, Atte Moilanen, Vuokko Heikinheimo, and Enrico Di Minin. 2018. "Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-based Experiences in Protected Areas." *Conservation Letters* 11 (1). Wiley Online Library: e12343.
- Hecht, Brent, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. "Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 237–46. ACM.
- Heikinheimo, Vuokko, Enrico Di Minin, Henriikki Tenkanen, Anna Hausmann, Joel Erkkonen, and Tuuli Toivonen. 2017. "User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey." *ISPRS International Journal of Geo-Information* 6 (3): 85. doi:10.3390/ijgi6030085.
- Hill, Linda L. 2009. *Georeferencing: The Geographic Associations of Information*. Cambridge, MA: MIT Press.
- Hollenstein, Livia, and Ross Purves. 2010. "Exploring Place through User-Generated Content: Using Flickr to Describe City Cores." *Journal of Spatial Information Science*. doi:10.5311/JOSIS.2010.1.3.
- Jones, C B, R S Purves, P D Clough, and H Joho. 2008. "Modelling Vague Places with Knowledge from the Web." *International Journal of Geographical Information Science* 22 (10): 1045–65.
- Keeler, Bonnie L, Spencer A Wood, Stephen Polasky, Catherine Kling, Christopher T Filstrup, and John A Downing. 2015. "Recreational Demand for Clean Water: Evidence from Geotagged Photographs by Visitors to Lakes." *Frontiers in Ecology and the Environment* 13 (2): 76–81.
- Korpilo, Silviya, Tarmo Virtanen, and Susanna Lehvävirta. 2017. "Smartphone GPS Tracking—Inexpensive and Efficient Data Collection on Recreational Movement." *Landscape and Urban Planning* 157. Elsevier: 608–17.
- Levin, Noam, Alex Mark Lechner, and Greg Brown. 2017. "An Evaluation of Crowdsourced Information for Assessing the Visitation and Perceived Importance of Protected Areas." *Applied Geography* 79 (February): 115–26. doi:10.1016/j.apgeog.2016.12.009.

- Li, Linna, Michael F Goodchild, and Bo Xu. 2013. "Spatial, Temporal, and Socioeconomic Patterns in the Use of Twitter and Flickr." *Cartography and Geographic Information Science* 40 (2). Taylor & Francis: 61–77.
- Manning, Christopher D, and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Vol. 999. Cambridge: MIT Press.
- Minin, Enrico Di, Henrikki Tenkanen, and Tuuli Toivonen. 2015. "Prospects and Challenges for Social Media Data in Conservation Science." *Frontiers in Environmental Science* 3 (September). doi:10.3389/fenvs.2015.00063.
- Moote, Margaret A, Mitchel P McClaran, and Donna K Chickering. 1997. "Theory in Practice: Applying Participatory Democracy Theory to Public Land Planning." *Environmental Management* 21 (6). Springer: 877–89.
- Norman, Patrick, and Catherine Marina Pickering. 2017. "Using Volunteered Geographic Information to Assess Park Visitation: Comparing Three on-Line Platforms." *Applied Geography* 89 (December): 163–72. doi:10.1016/j.apgeog.2017.11.001.
- Orsi, Francesco, and Davide Geneletti. 2013. "Using Geotagged Photographs and GIS Analysis to Estimate Visitor Flows in Natural Areas." *Journal for Nature Conservation* 21 (5). Elsevier: 359–68.
- Oteros-Rozas, Elisa, Berta Martín-López, Nora Fagerholm, Claudia Bieling, and Tobias Plieninger. 2017. "Using Social Media Photos to Explore the Relation between Cultural Ecosystem Services and Landscape Features across Five European Sites." *Ecological Indicators*, February. doi:10.1016/j.ecolind.2017.02.009.
- Purves, Ross S, Alistair Edwardes, and Joe Wood. 2011. "Describing Place through User Generated Content." *First Monday* 16 (9). <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3710/303>.
- Purves, Ross S, and William A Mackaness. 2016. "A Methodological Toolbox for Exploring Collections of Textually Annotated Georeferenced Photographs." In *European Handbook of Crowdsourced Geographic Information*, edited by C Capineri, M Haklay, H Huang, V Antoniou, J Kettunen, F Ostermann, and R Purves, 145. London: Ubiquity Press.
- Sessions, Carrie, Spencer A. Wood, Sergey Rabotyagov, and David M. Fisher. 2016. "Measuring Recreational Visitation at U.S. National Parks with Crowd-Sourced Photographs." *Journal of Environmental Management* 183 (December): 703–11. doi:10.1016/j.jenvman.2016.09.018.
- Sonter, Laura J., Keri B. Watson, Spencer A. Wood, and Taylor H. Ricketts. 2016. "Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media." Edited by Jian Yang. *PLOS ONE* 11 (9): e0162372. doi:10.1371/journal.pone.0162372.
- Spielman, Seth E. 2014. "Spatial Collective Intelligence? Credibility, Accuracy, and Volunteered Geographic Information." *Cartography and Geographic Information Science* 41 (2). Taylor & Francis: 115–24.
- Tenerelli, Patrizia, Urška Demšar, and Sandra Luque. 2016. "Crowdsourcing Indicators for Cultural Ecosystem Services: A Geographically Weighted Approach for Mountain Landscapes." *Ecological Indicators* 64 (May). Elsevier: 237–48.

doi:10.1016/j.ecolind.2015.12.042.

- Tenkanen, Henrikki, Enrico Di Minin, Vuokko Heikinheimo, Anna Hausmann, Marna Herbst, Liisa Kajala, and Tuuli Toivonen. 2017. "Instagram, Flickr, or Twitter: Assessing the Usability of Social Media Data for Visitor Monitoring in Protected Areas." *Scientific Reports* 7 (1): 17615. doi:10.1038/s41598-017-18007-4.
- Tieskens, Koen F., Catharina J.E. Schulp, Christian Levers, Juraj Lieskovský, Tobias Kuemmerle, Tobias Plieninger, and Peter H. Verburg. 2017. "Characterizing European Cultural Landscapes: Accounting for Structure, Management Intensity and Value of Agricultural and Forest Landscapes." *Land Use Policy* 62 (March): 29–39. doi:10.1016/j.landusepol.2016.12.001.
- Tufekci, Zeynep. 2014. "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls." *ICWSM* 14: 505–14.
- Tversky, Barbara, and Kathleen Hemenway. 1983. "Categories of Environmental Scenes." *Cognitive Psychology* 15 (1): 121–49.
- Wardman, Mark. 1988. "A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour." *Journal of Transport Economics and Policy*. JSTOR, 71–91.
- Wartmann, F M, E Acheson, and R S Purves. 2018. "Describing and Comparing Landscapes Using Tags, Texts, and Free Lists: An Interdisciplinary Approach." *International Journal of Geographical Information Science*, April, 1–21. doi:10.1080/13658816.2018.1445257.
- Wartmann, F M, E Egorova, C Derungs, D M Mark, and R S Purves. 2015. "More Than a List: What Outdoor Free Listings of Landscape Categories Reveal About Commonsense Geographic Concepts and Memory Search Strategies." In *Spatial Information Theory. Lecture Notes in Computer Science*, edited by S I Fabrikant, Martin Raubal, S Freundschuh, and Scott Bell, 9368:224–43. Heidelberg: Springer. doi:10.1007/978-3-319-23374-1_11.
- Wartmann, F M, and R S Purves. 2018. "Investigating Sense of Place as a Cultural Ecosystem Service in Different Landscapes through the Lens of Language." *Landscape and Urban Planning* 175 (July): 169–83. doi:10.1016/j.landurbplan.2018.03.021.
- Wood, Spencer A, Anne D Guerry, Jessica M Silver, and Martin Lacayo. 2013. "Using Social Media to Quantify Nature-Based Tourism and Recreation." *Scientific Reports* 3. Nature Publishing Group.
- Zanten, Boris T. van, Derek B. Van Berkel, Ross K. Meentemeyer, Jordan W. Smith, Koen F. Tieskens, and Peter H. Verburg. 2016. "Continental-Scale Quantification of Landscape Values Using Social Media Data." *Proceedings of the National Academy of Sciences* 113 (46): 12974–79. doi:10.1073/pnas.1614158113.

Anhang A Übersicht zu Studien über Besucherverhalten und – wahrnehmung mit Datengrundlage soziale Medien

Study	Geographic area	Study area size	Environment	Information on visitor frequency	Information on behaviour	Perception and values	Type of data	Analysis method
Barry, S. J. (2014). Using social media to discover public values, interests, and perceptions about cattle grazing on park lands. <i>Environmental Management</i> , 53(2), 454–464.	San Francisco Bay area California	540km ²	rangelands	No	qualitative	Public perception of gattle grazing and cows on public lands	Georeferenced Flickr fotos	Selection of geotagged Flickr images by keywords in study area Content analysis through manual annotation of photo content and comments into categories
Campelo, M. B., & Nogueira Mendes, R. M. (2016). Comparing webshare services to assess mountain bike use in protected areas. <i>Journal of Outdoor Recreation and Tourism</i> .	Sintra-Cascais Natural Park, Portugal	145km ²	various	No	Spatial behaviour of mountain bike users, intensity of use in particular hotspot areas of the natural area, identified illegal trail use	Not assessed	Mountain biking tracks from GPSies and Wikiloc.com	Calculate track density for raster cells as intensity of use Classifying use of trails as official and inofficial/illegal trail use
Capineri, C. (2016). Kilburn High Road Revisited. <i>Urban Planning</i> , 1(2), 128.	Kilburn road, London, UK	Approx. 2km ²	Urban	No	Activities people associate with a place	location, locale and sense of place	Georeferenced Tweets and Flickr fotos	Semantics, content analysis and frequency of word counts
Casalegno, S., Inger, R., DeSilvey, C., & Gaston, K. J. (2013). Spatial Covariance between Aesthetic Value & Other Ecosystem Services. <i>PLoS ONE</i> , 8(6), e68437.	Cornwall, UK	3,563 km ²	mixed	yes, as a proxy of aesthetic value	no	Aesthetic value of places	Georeferenced Panoramio images	Counting number of users per 1km ² and linking it to population density and coastal/non-coastal locations
Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. <i>Landscape and Urban Planning</i> , 170, 283–292.	New Brunswick and British Colombia		Proposed hydroelectric dams			Perception and values associated with water bodies: aesthetics, sense of home, community attachment, cultural identity, lifestyle, memory	Instagram	Automated text analysis through Netlytic, Semi-inductive content analysis of captions and mapping of landscape values through kernel density estimation
Chesnokova, O., Nowak, M., & Purves, R. S. (2017). A crowdsourced model of landscape preference. In Iseco Clementini, C. Donnelly, M. Yuan, C. Kray, P. Fogliaroni, & A. Ballatore (Eds.), <i>LIPics-Leibniz International Proceedings in Informatics</i> (Vol. 86, p. 19:1-19:13). https://doi.org/10.4230/LIPics.COS.IT.2017.19								
Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. Vande. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. <i>Tourism Management</i> , 57, 295–310.	Cilento, Italy	4.9 km ²	Mixed cultural landscapes	Identifying number of visitors per country of origin and domestic/foreign tourist Season patterns in visitation frequency	Number of days spent visiting Italy/Cilanto Movement trajectories of tourists, differences in travel behavior between domestic and foreign tourists Location of popular tourist sites		Geotagged tweets	Trajectory mining from geolocated tweets
Donahue, M. L., Keeler, B. L., Wood, S. A., Fisher, D. M., Hamstead, Z. A., & McPhearson, T. (2018). Using social media to understand drivers of urban park visitation in the Twin Cities, MN. <i>Landscape and Urban Planning</i> , 175, 1–10.	Twin Cities, Minnesota, USA	Various small city parks and urban green areas	Urban green spaces	Correlation with official visitor counts both for flickr and twitter data	water features, number of amenities, length of trails, and population density of nearby neighborhoods had positive and significant effects on visitation rate	Not assessed	Georeferenced Flickr photos and Tweets	

Study	Geographic area	Study area size	Environment	Information on visitor frequency	Information on behaviour	Perception and values	Type of data	Analysis method
Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. <i>Landscape and Urban Planning</i> , 142, 173–186.	Various (e.g. Yosemite NP, State of California, Germany)	Various (from National Park to country size)	Mixed, including cityscapes and protected areas	Counting number of photographs uploaded corrected for bulk-uploads and multiple pictures per 10m ²	Spatio-temporal behavior through density distribution of content Analysing spatial behavior through timestamps in content	Characterising places based on tags	Georeferenced Flickr images	Spatio-temporal tag maps
Figuerola-Alfaro, R. W., & Tang, Z. (2017). Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. <i>Journal of Environmental Planning and Management</i> , 60(2), 266–281.	Nebraska	200,356 km ²	mixed	Not assessed	Not assessed	State parks and national parks had the highest density of photos with nature related tags per area in Nebraska	Georeferenced images from Panoramio and Flickr	Selection of photos with pre-defined tags relating to cultural and natural aspects of the environment Clustering and hot/coldspot analysis
Guerrero, P., Möller, M.S., Olafsson, A.S., Snizek, B., 2016. Revealing Cultural Ecosystem Services through Instagram Images: The Potential of Social Media Volunteered Geographic Information for Urban Green Infrastructure Planning and Governance. <i>Urban Plan.</i> 1, 1.	Copenhagen, Denmark	179 km ²	Urban	Not assessed	Where people upload images related to nature	Perception of nature in urban areas	Georeferenced Instagram images with hashtag #sharingcph	Manual annotation of image content into pre-defined categories Cluster and distance analysis
Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in Protected Areas. <i>Conservation Letters</i> , 11(1), e12343.	Kruger National Park, South Africa	19,485 km ²	savanna	Not assessed	Assessed through survey and content of posted images	Preferences expressed on social media (for animals and landscapes) was similar to preferences for megafauna expressed in survey	Georeferenced Flickr and Instagram images	Assessing proportion of visitors using social media platforms (80%) through a survey: Use of Flickr as social media platform was not significantly correlated with age Social media use decreased with income
Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. <i>ISPRS International Journal of Geo-Information</i> , 6(3), 85.	Pallas-Yllästunturi National Park, Finland	1,020 km ²	Mix of mountainous areas with mosaic of natural peat-bogs and forests	Correlates well with official counts for highly frequented sites but less for data sparse areas	Activities conducted by visitors and temporal aspects of visitors' recreational behaviours	Not assessed	Georeferenced Instagram images (with predefined location)	Selection of Instagram images from pre-defined points of interests within national park Manual classification of ~20'000 images to infer activities
Keeler, B. L., Wood, S. A., Polasky, S., Kling, C., Filstrup, C. T., & Downing, J. A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. <i>Frontiers in Ecology and the Environment</i> , 13(2), 76–81.	Minnesota and Iowa, USA	n/a	oligotrophic and eutrophic lakes	Estimating unique photo-user days as a proxy for visitation	Estimating time and distance travelled between picture locations and a user's home location	Number of visits were significantly correlated with water clarity. Water clarity also correlated positively with travel time (revealed preference for water quality)	Georeferenced Flickr images	Calculating average person-user-days and comparing them with visitor statistics Using home location provided by Flickr users to calculate distances travelled by domestic recreational users
Korpilo, S., Virtanen, T., & Lehtävirta, S. (2017). Smartphone GPS tracking—Inexpensive and efficient data collection on recreational movement. <i>Landscape and Urban Planning</i> , 157, 608–617.	Keskuspuisto park, Helsinki, Finland	11ha	Urban park, urban forest	Not assessed	Collect GPS tracks from users: 22 mountain bike tracks, 25 running tracks. Identify recreational movement and use intensities of mountain bikers and runners as well as off-trail behavior: Mountain bikers go more off trail than runners	Not assessed	Asking participants to upload their tracking data collected on their own devices and uploaded on various platforms (Sports Tracker, Strava, Endomondo, Polar)	Distinguishing on trail and off-trail movements by intersecting GPS tracks from users with buffer around official trail network

Study	Geographic area	Study area size	Environment	Information on visitor frequency	Information on behaviour	Perception and values	Type of data	Analysis method
Pastur, G. M., Peri, P. L., Lencinas, M. V., García-Lorente, M., & Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. <i>Landscape Ecology</i> , 31(2), 383–399.	Santa Cruz and Tierra del Fuego province, Southern Patagonia, Argentina	Santa Cruz 243,900 km ² Tierra del Fuego 21,300 km ²	Mixed, mainly cold arid steppe	Not assessed	Not assessed	Main CES identified was aesthetic value (81%). All other identified values were marginal, e.g. recreation 3.7% Spatial patterns for different CES were different. Pictures with aesthetic value were taken near water bodies.	Geotagged Panoramio images	Assessing image content and counting number of images to assess cultural ecosystem services of (i) aesthetic value, (ii) existence value, (iii) recreation, and (iv) local identity
Norman, P., & Pickering, C. M. (2017). Using volunteered geographic information to assess park visitation: Comparing three on-line platforms. <i>Applied Geography</i> , 89, 163–172.	Protected areas in Queensland, Australia	5.71km ² – 205.9km ²	peri-urban and rural protected areas	Average monthly trail counts correlate with users on social media (R ² = 0.803) but are different orders of magnitude	Assess visitor activities (biking vs. walking) based on fitness logs Relative popularity of trails Spatial use patterns (illegal use of sacred sites that are off limit to the public)	Not assessed directly	Fitness apps: MapMyFitness GPSies Wikiloc	Calculating line density Comparison of users logging routes on fitness apps with physically installed trail counters
Orsi, F., & Geneletti, D. (2013). Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. <i>Journal for Nature Conservation</i> , 21(5), 359–368.	Dolomites UNESCO world heritage site, Italy	1490 km ²	Mountain landscapes	density flow model that uses popular sites identified from geotagged images and slope models to estimate visitor numbers on trails model calibrated with actual visitor counts predicts visitor flows in same number of magnitude as actual counts (r=0.88)	Flow model can predict behavior that less steep trails are more popular than steep trails	Not assessed	Geotagged Panoramio photos	Identifying access points and popular tourist spots for a gravitational model of visitor flows by calculating high densities of georeferenced photographs
Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2017). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. <i>Ecological Indicators</i> .	One landscape each in Estonia, Greece, Spain, Sweden, Switzerland	47km ² - 334km ²	Boreal, Alpine, and Mediterranean	Not assessed	Identifying visitor activities through manual annotation of image content	CES: The most depicted cultural ecosystem service was recreation Flickr contained more depictions of CES than Panoramio despite lower N Landscape elements: Water bodies are highly preferred in images Forests in Switzerland and Sweden were overrepresented in images compared to land cover	Geotagged Flickr and Panoramio images	Manual selection of images containing landscapes Manual annotation of content of landscape images to identify cultural ecosystem services (e.g. cultural heritage, recreation) and landscape elements (agriculture, mountain, trees, grasslands, sea, lake) Assessing landscape elements in comparison to actual land cover to assess relative preference for certain landscape elements
Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. <i>Ecological Indicators</i> , 53, 187–195.	Singapore	0.03-1.33km ²	Mangrove forests	Not assessed	Linking number of images to properties of the areas: distance from footpaths was most important explanatory factor for number of images	Content of images reveals preferences and cultural ecosystem services consumed (nature-based recreation, social recreation), which were compared between different areas Lack of photographs (e.g. in coastal areas) can be used to guide management and information provision to visitors	Geotagged Flickr photographs	Manual content annotation of Flickr images for different CES categories and comparing between 4 different study sites Modelling the probability of presence of different types of CES in photographs using Maximum Entropy model
Richards, D. R., & Tunçer, B. (2017). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. <i>Ecosystem Services</i> .	Singapore	719.9 km ²	mixed	Not assessed	Not assessed	Images with nature-related content were clustered around nature attractions and protected areas	Geotagged Flickr photographs	Using automated image recognition to count Flickr images containing nature and animals to assess cultural ecosystem services

Study	Geographic area	Study area size	Environment	Information on visitor frequency	Information on behaviour	Perception and values	Type of data	Analysis method
Sonter, L. J., Watson, K. B., Wood, S. A., & Ricketts, T. H. (2016). Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media. <i>PLOS ONE</i> , 11(9), e0162372.	Vermont, USA	State of Vermont, protected areas: 5800km ²	mixed	Comparing counts of Flickr images with survey data: positive correlation between visitor counts and social media uploads	Not assessed	Not assessed	Geotagged Flickr photographs	Counting number of individual users per day in defined grid square
Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. <i>Ecological Indicators</i> , 64, 237–248.	Quatre Montagnes region in Vercors Natural Regional Park, France	Ca. 50 km ²	Forest and grassland habitats	Not assessed	Not assessed	Explaining number of photos as a proxy for revealed preferences in gridcells through geographic weighted regression with explanatory variables including distance to trails, presence of water bodies and number of potential visitors. High residuals in the model were explained by presence of special landscape features that attract visitors.	Georeferenced images manually filtered for image content relating to CES	Using photo counts as dependent variable and landscape settings as explanatory variables to explain spatial variability in cultural ecosystem service consumption
Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. <i>Scientific Reports</i> , 7(1), 17615.	21 national parks in South Africa 35 parks in Finland	mixed	mixed	Social media data correlated with the monthly visitation patterns better in more visited parks and in parks with higher quantity of social media data Estimates were different by orders of magnitude	Not assessed	Not assessed	Georeferenced data from Flickr, Twitter and Instagram	Calculating active social media users per day
Tieskens, K. F., Schulp, C. J. E., Levers, C., Lieskovský, J., Kuemmerle, T., Plieninger, T., & Verburg, P. H. (2017). Characterizing European cultural landscapes: Accounting for structure, management intensity and value of agricultural and forest landscapes. <i>Land Use Policy</i> , 62, 29–39.	Europe (EU27 + Switzerland)	4 413 652 km ²	mixed	Not assessed	Not assessed	Areas in Europe that were identified as being typical cultural landscapes were Tuscany, the Dehesas or the bocage in Normandy	Panoramio pictures	Counting number of pictures per gridcell and capping at 10 pictures / cell. Using counts of images on Panoramio as part of indicator for landscape meaning (together with geographic food labels as number of products of designated origins (PDOs))
Van Berkel, D. B., Tabrizian, P., Dorning, M. A., Smart, L., Newcomb, D., Mehaffey, M., ... Meentemeyer, R. K. (in press). Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. <i>Ecosystem Services</i> .	North Carolina, Middle Atlantic Coastal Plain	2251.7 km ²	Agriculture and coastal ecosystems	Not assessed	Not assessed	Content analysis: Large proportion of landscape images contain water bodies or beach. Regression model: ocean, beaches, protected areas and trails were positively associated with viewshed intensity. Forests and agricultural areas were negatively associated with viewshed intensity	Georeferenced images from Panoramio	For Panoramio foto locations in study site (n=16,582). Analysed image content and calculated viewsheds based on locations. Calculated cumulative viewsheds as an indication of highly photographed landscape features. For highest cumulative viewsheds, take random sample of 1% of raster cells (equaling a total of square kilometers) fit a regression model

Study	Geographic area	Study area size	Environment	Information on visitor frequency	Information on behaviour	Perception and values	Type of data	Analysis method
Walden-Schreiner, C., Leung, Y.-F., & Tateosian, L. (2018). Digital footprints: Incorporating crowdsourced geographic information for protected area management. <i>Applied Geography</i> , 90, 44–54.	Hawaii Volcanoes national park, Hawaii, USA	1300km ²	Volcanic landscape	MaxEnt models show visitor hotspots within national park Hotspots were located along roads and trails near visitor center at near attractions	Visitors use accessible areas along roads and near visitor attractions	Not assessed	Georeferenced flickr images	MaxEnt modelling of visitor locations based on Flickr images
Wartmann, F. M., Acheson, E., & Purves, R. S. (2018). Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach. <i>International Journal of Geographical Information Science</i> , 1–21.	10 different study sites in Switzerland	n/a	river, hill, mountain, forest, moor landscapes	Not assessed	Visitors stated mostly hiking as activities in investigated blogs	Characterising different landscape types based on public perception using landscape terms extracted from social media	Georeferenced Flickr images and hiking blogs combined with interviews	Georeferencing hiking blogs using place names used in descriptions Generating footprints for hiking blogs for each study site to download Flickr data Characterising landscapes based on text extracted from social media data
Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. <i>Scientific Reports</i> , 3.	Global (836 recreational sites from Disneyland to Uganda NP)	Recreational sites (80m ² of US National Memorial to 30448 km ² Arctic NP)	Mixed (natural and cultural attractions)	Yes	Not assessed	Not assessed	Georeferenced flickr images	Analysis of person-days Correlation with other variables