Selection of a network of large lakes and reservoirs suitable for global environmental change analysis using Earth Observation
Politi, E.; MacCallum, S.; Cutler, M. E. J.; Merchant, C. J.; Rowan, J. S.; Dawson, T. P

Published in:
International Journal of Remote Sensing

DOI:
10.1080/01431161.2016.1192702

Publication date:
2016

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):
Selection of a network of large lakes and reservoirs suitable for global environmental change analysis using Earth Observation

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>TRES-PAP-2015-0910.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>IJRS Research Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>08-Mar-2016</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Politi, Eirini; University College Cork National University of Ireland, Environmental Research Institute MacCallum, Stuart; University of Edinburgh, School of Geosciences Cutler, Mark; University of Dundee, School of Social Sciences Merchant, Christopher; University of Reading, Department of Meteorology Rowan, John; University of Dundee, School of Social Sciences Dawson, Terence; University of Dundee, School of Social Sciences</td>
</tr>
<tr>
<td>Keywords:</td>
<td>ATSR, lakes</td>
</tr>
<tr>
<td>Keywords (user defined):</td>
<td>environmental change, site selection, GloboLakes</td>
</tr>
</tbody>
</table>
List of Authors

Eirini Politi (corresponding author), MaREI Centre, Environmental Research Institute, University College Cork, Beaufort Building, Ringaskiddy, Co. Cork, Ireland
Tel: +353 21 486 4339
Email: eirini.politi@ucc.ie

Stuart MacCallum, School of Geosciences, University of Edinburgh, Edinburgh EH9 3FE, United Kingdom
Tel: N/A
Email: stuart.maccallum@astrosat.biz

Mark E.J. Cutler, School of Social Sciences, University of Dundee, Dundee DD1 4HN, United Kingdom
Tel: +44 1382 385446
Email: m.e.j.cutler@dundee.ac.uk

Chris J. Merchant, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, United Kingdom
Tel: +44 118 378 7335
Email: c.j.merchant@reading.ac.uk

John S. Rowan, School of Social Sciences, University of Dundee, Dundee DD1 4HN, United Kingdom
Tel: +44 1382 384024
Email: j.s.rowan@dundee.ac.uk

Terry P. Dawson, School of Social Sciences, University of Dundee, Dundee DD1 4HN, United Kingdom
Tel: +44 1382 388064
Email: t.p.dawson@dundee.ac.uk
Disclosure Statement

The authors of this manuscript own copyright and wish to assign copyright to Taylor & Francis by means of a publishing agreement should this manuscript is accepted and moves into production. We have seen, read, and understood the Taylor & Francis guidelines on copyright.
Selection of a network of large lakes and reservoirs suitable for global environmental change analysis using Earth Observation

E. Politi\textsuperscript{1,a}, S. MacCallum\textsuperscript{b}, M.E.J. Cutler\textsuperscript{c}, C.J. Merchant\textsuperscript{d}, J.S. Rowan\textsuperscript{c} and T.P. Dawson\textsuperscript{c}

\textsuperscript{a} MaREI Centre, Environmental Research Institute, University College Cork, Beaufort Building, Ringaskiddy, Co. Cork, Ireland
\textsuperscript{b} School of Geosciences, University of Edinburgh, Edinburgh EH9 3FE, United Kingdom
\textsuperscript{c} School of Social Sciences, University of Dundee, Dundee DD1 4HN, United Kingdom
\textsuperscript{d} Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, United Kingdom

Abstract

The GloboLakes project, a global observatory of lake responses to environmental change, aims to exploit current satellite missions and long remote sensing archives to synoptically study multiple lake ecosystems, assess their current condition, reconstruct past trends to system trajectories and assess lake sensitivity to multiple drivers of change. Here we describe the selection protocol for including lakes in the global observatory based upon remote sensing techniques and an initial pool of the largest 3,721 lakes and reservoirs in the world, as listed in the Global Lakes and Wetlands Database. An 18-year long archive of satellite data was used to create spatial and temporal filters for the identification of water bodies that are appropriate for remote sensing methods. Further criteria were applied and tested to ensure the candidate sites span a wide range of ecological settings and characteristics; a total 960 lakes, lagoons and reservoirs were selected. The methodology proposed here is applicable to new generation satellites, such as the European Space Agency Sentinel-series.

Keywords: GloboLakes, (A)ATSR, environmental change, lakes, site selection, global

\textsuperscript{1} Corresponding author: eirini.politi@ucc.ie
1. Introduction

The importance of lake ecosystems as regulators and sentinels of environmental change is well recognised (e.g. Adrian et al. 2009; Schindler 2009; Vincent 2009; Williamson et al. 2009, 2014). In order to assess what factors control the sensitivity and susceptibility of lakes to environmental change, it is necessary to adequately characterise and represent the range of lake responses that occur across the globe. This requires a comprehensive set of study lakes that span a range of lake types and ecological settings to ensure enough representatives from the large and diverse global lake population are included. According to recent estimates, the global population of freshwater bodies with a surface area exceeding 0.002 km\(^2\) (0.2 ha) is around 117 million (Verpoorter et al. 2014), which raises practical difficulties for the simultaneous and continuous monitoring of freshwater at global scales. Conventional field-based and in situ sampling are time consuming and labour intensive, meaning that a tiny fraction (< 0.003%) of the global population of lakes are systematically monitored. Studies on the impact of environmental (including climate) change require analysis of trends and change over long time scales (e.g. decadal), as well as across a range of spatial scales. Tools such as remote sensing offer the potential to provide the data for such analysis, but in limnological research a combination of different types of observations are often required and at relatively higher resolutions than in other domains. Despite this, a number of studies have made use of existing satellite-based sensors (some of them with data spanning 35-year long archives) to monitor change in lake water properties with very promising results (e.g. MacCallum and Merchant 2012; Politi et al. 2012; Palmer et al. 2015a).

Previous limnological studies that have looked at coherency and spatiotemporal patterns of lake response to environmental change are geographically limited to between local to sub-continental scales. Meteorological variables at local and regional scales and synoptic scale climatic phenomena have been shown to strongly influence various lake attributes, such as the thermal regime (e.g. Livingstone and Dokulil 2001; Arhonditsis et al. 2004; Wilhelm et al. 2006; Golosov et al. 2012; Read et al. 2014), ice phenology (Latifovic and Pouliot 2007), planktonic communities (e.g. Weyhenmeyer, Blenckner and Pettersson 1999; Gerten and Adrian 2000; Weyhenmeyer et al. 2002) and nutrient concentrations (Blenckner et al. 2007; Jones and Brett 2014). However, climatic changes are predicted to occur non-uniformly across the globe (Hardy 2003) and will have different effects on lakes depending on their geographic location (Adrian et al. 2009). This is likely to lead to complex response patterns requiring comprehensive understanding of other factors that complicate the linkages between lake water quality and climate change, including catchment land management changes, linked agricultural policies, point and diffuse pollution, habitat destruction and the impacts of invasive alien species (UNEP 2000; EUROPA WFD 2011). To the best knowledge of the authors, there has been no study that aims to understand lake behavioural response to environmental change across the globe by studying the linkages between hundreds of lakes and their catchments, and accounting for both human-induced and naturally occurring environmental change.

The United Kingdom (UK) GloboLakes project is using remote sensing-based approaches to establish a global observatory of the world’s largest lakes, to reveal their responses (individually and collectively) to climatic and land-based drivers of change over recent decades. Similar (but smaller scale) projects of this kind have developed
their own protocols for selecting sample lakes, including approaches such as (a) ‘judgemental site selection’ (e.g. NOAA National Status and Trends programme (NOAA NS&T 2016), US Geological Survey (USGS) National Water-Quality Assessment programme (USGS NAWQA 2016)), where sites are selected based on their representativeness of regional characteristics; or (b) statistical methods (e.g. US Forest Service (USFS) Forest Inventory and Analysis programme (USFS FIA 2016), US EPA (Environmental Protection Agency) Environmental Monitoring & Assessment Program (US EPA EMAP 2016)), where the whole US lake population was sampled in a probabilistic manner (Stevens 1994). However, neither approach accounts for variability in lake morphology, inclusive of attributes such as size, presence of islands or irregularity of the shoreline, and hence the suitability of the system to be monitored using remote sensing at spatial resolutions relevant to archived and current systematically acquired data.

To detect changes in the spatial patterns and coherency of lake response to climate change requires information on key parameters of water quality (e.g. temperature, chlorophyll and transparency) and quantity (e.g. water level and volume) over different seasons and years from a global set of lakes at a high temporal resolution. Whilst data from new sensors, such as ESA’s Sentinel series of satellites, provide opportunities for regular monitoring going forward, we aim to exploit the existing archive of remotely sensed observations to detect change in over the past two decades. Observations from satellite sensors, such as the Advanced Along Track Scanning Radiometer (AATSR) and the Medium Resolution Imaging Spectrometer (MERIS) on-board the ENVISAT satellite provide long archives of imagery with sufficient frequency of observation (potential view of a given lake every 1 to 3 days compared to 16 days for Landsat satellites, for example) enabling changes in lake phenology to be investigated and mapped over multiple years (e.g. Palmer et al. 2015b; ESA ArcLake 2016). In addition, radar altimeters on various satellites (e.g. TOPEX/POSEIDON, Jason-1, Jason-2/OSTM and ENVISAT) have been used to provide information on lake/reservoir water quantity and freshwater level fluctuations (Politi et al. 2016). However, whilst providing a high temporal resolution, the spatial resolution of the abovementioned thermal and optical instruments (at best, 1000 m and 300 m for AATSR and MERIS respectively) impose limitations on the number of lakes that can be studied with currently available remote sensing systems, with lake size and shape being particularly important characteristics in determining whether a lake can be reliably observed. In this sense, properties such as minimum detectable size of a lake and/or the presence of islands that may contaminate the pure water leaving radiance detected by the satellite instrument, need to be taken into account. Therefore, in order to exploit the information available in the archive of remotely sensed observations, a series of selection criteria were needed to determine a sample of lakes that not only adhere to the principle of lake diversity and differential response to environmental change, but also take account of impositions placed upon data collection at spatial resolutions of suitable satellite archives and currently available sensors. It is anticipated that the results of this analysis will show both the potential and limitations of exploiting the current satellite observation archive.

Due to the long archive of data available, the temporal resolution required to potentially capture seasonal lake phenology and the fidelity of temperature retrieval, we based our site selection methodology on a spatially coarse sensor, namely the ENVISAT AATSR (2002-2012) and its predecessor the ERS-2 Along-Track Scanning
Radiometer (ATSR2) (1995-2003) to select lakes based on the ‘worst case’ scenario of
spatial remote sensing data, i.e. lakes that can be reliably remotely sensed with AATSR
will more than likely be suitable for study with MERIS and finer resolution data.

This paper describes the procedure adopted by the GloboLakes project to select
water bodies across the globe that are suitable for analysis by remote sensing
techniques, constrained by the limitations of spatial resolution described above. Lake
selection was based on minimum detectable lake size, average number of daytime
observations per year, and size of detectable area in respect to total lake size. For this
purpose, the Global Lakes and Wetlands Database (GLWD) Level-1 inventory (Lehner
and Döll 2004) was used as an initial pool of lakes, from which candidate lakes were
selected. Additional criteria were applied and tested to ensure the final list included
water bodies that spanned a wide range of ecological characteristics and were situated
within all biomes of the world. The methodology proposed here can be applied to any
limnological study at local, regional or global scales, and provides an important insight
into the opportunities and limitations of systematic observations of lakes, afforded from
both archive data and future missions.

2. Definitions

Apart from natural lakes (both permanent and ephemeral) and artificial lakes
(reservoirs), GLWD also contains coastal lagoons and other water bodies identified in
this study as coastal bays, rivers, estuaries, fjords and glaciers. The definitions of these
terms as adopted for the GloboLakes project are given below.

Natural lakes
Static (lentic) bodies of water occupying inland basins (Herdendorf 1982) without
direct connection to the sea (Lehner and Döll 2004) that are discrete and may contain a
number of basins between which there is generally multi-directional exchange and
mixing of water at least during part of the year (Nixon, Grath and Bøgestrand 1998).
They can be of either permanent or seasonal character.

Ephemeral lakes
Natural lakes, whose water level varies significantly within a year or between years
and may dry out partially or completely, depending on various reasons (e.g. seasonality,
climatic variability, human pressure, etc.). These systems are common in arid regions,
but are also found in other parts of the world, and in their dry states can be major
sources of atmospheric mineral aerosols (dust) (e.g. Mahowald et al 2003), which can
seriously impact the regional climate, ecosystems and human health (Rashki et al 2013).

Coastal lagoons
Static bodies of water separated from the oceans by spits and barrier bars
(Herdendorf 1982). According to a classification by Kjerfve (1986), there are three
types of coastal lagoons based on the number of entrance channels (inlets) and, thus, the
degree of water exchange with the ocean: (a) choked, (b) restricted, and (c) leaky.

Choked lagoons have only one inlet, which restricts the influence of tidal currents and
water level fluctuations in the lagoon. They can be either parallel to the shore or, when
associated with river deltas, at a right angle to the shore. Restricted lagoons have two or
more inlets and a well-defined tidal circulation, whilst leaky lagoons exhibit numerous
inlets and are the most influenced by tidal currents in all three lagoon types. Both restricted and leaky are most usually oriented parallel to the shore.

Reservoirs

Are artificial water bodies, in this case dammed river valleys, or natural lakes deepened and extended by outflow control (impounding dam or controlling sluice), typically with pronounced water level controls. The nature of the water level regulation may be modest or dramatic depending on the purpose e.g. water supply, hydropower, navigation etc. and environmental context (cf. Herdendorf 1982; Lehner and Döll 2004).

Coastal bay

A body of water connected to an ocean or sea by a broad opening, where the land curves inwards. Bays also exist as inlets to any larger body of water such as lakes, ponds and estuaries.

River

A natural stream of water flowing in a channel to the sea, a lake, or another river.

Estuary

The wide mouth of a river where the river meets the sea/tide.

Fjord

A long, narrow, deep inlet of the sea between high cliffs, typically formed by glacial erosion.

Glacier

A slowly moving mass or river of ice formed by the accumulation and compaction of snow on mountains or near the poles.

3. Remote sensing instruments

The European Space Agency (ESA) ERS-2 ATSR2 and ENVISAT AATSR were used in this study. The choice of instrument was based on two criteria: (a) it provides a long archive of lake water surface temperature data that will be used in GloboLakes, and (b) has a much coarser resolution (1 km) than the optical instruments used to map other lake water quality parameters (e.g. biological and chemical), which means the size and shape of lakes appropriate for use with the ATSR2/AATSR data will also be suitable for other optical and thermal sensors (e.g. ENVISAT MERIS (300 m) and Copernicus Sentinel-2 Multispectral Imager (MSI; 10-20 m), Sentinel-3 Ocean and Land Colour Instrument (OLCI; 300 m) and Sea and Land Surface Temperature Radiometer (SLSTR; 500-1000m)).

The ATSR2/AATSR were space-borne instruments flown at an altitude of 800 km and were operational in the periods 1995-2003 and 2002-2012, respectively. ATSR2 and AATSR were dual-view, multi-channel, imaging radiometers with 512 km swath widths, revisit times of three days over the tropics, with more frequent observation possible at high latitudes. Both featured spatial resolutions of 1 km² at nadir, with approximately 3 km² in the forward view, and had stable late-morning orbits (10:00 h or 10:30 h local equator crossing time with minimal drift), yielding consistent overlap.
periods to support their application to global climate monitoring (MacCallum and Merchant 2012).

4. Identification of GloboLakes Study Sites

The identification of GloboLakes study sites from the original GLWD Level-1 inventory was based upon the application of selection criteria that satisfied two objectives. Firstly, the criteria were designed to ensure a candidate lake list complied with the restrictions imposed by remote sensing in terms of the morphology of lakes that can be detected, as outlined above. Secondly, selection was guided by the need to maximise the range of physical lake types and ecological behaviours spanning the continuum of lake landscape settings across the world’s major biomes. Prior to the application of the selection criteria, a quality assurance check was performed on the GLWD Level-1 inventory to ensure the data were ‘fit for purpose’ (see Section 4.1).

The site selection process is presented in Fig. 1.

Insert Figure 1 near here

4.1 GLWD Data Quality Assurance

The name, location and type of all GLWD Level-1 lakes was validated using Google Maps and Google Earth (imagery dates 2014-2015) to ensure that all listed water bodies were still in existence and that any incorrect input data (associated with location, extent and spatial representation) were identified.

On investigating the database it became apparent that a small number of lakes were represented by a circle of approximate lake size instead of the actual lake outline (examples are shown in Fig. 2). These were excluded from further analysis, as whilst it would be possible to produce a map of water bodies from the water mask algorithm applied here (MacCallum and Merchant 2012), this was not the aim of this exercise and our objective was to use existing lake outline information instead of generating new outlines or correcting errors (although this may be the focus of additional work in the future). As a result of this, 51 water bodies were excluded reducing the original GLWD Level-1 inventory to 3,670 water bodies. Additionally, using information on the location of dams and reservoirs from the GRanD database (Lehner et al. 2011), the GLWD pre-determined status of ‘lake’ or ‘reservoir’ was validated for all water bodies. As a result, some GLWD ‘lakes’ were assigned a ‘reservoir’ status. Water bodies that were not lakes or reservoirs, were assigned a new status. The terms ‘lagoon’, ‘coastal bay’, ‘river’, ‘estuary’, ‘fjord’, ‘glacier’ (defined in Section 2) and ‘other’ were used to fulfil this purpose.

Finally, almost 60% of the lake names were missing from the original GLWD Level-1 inventory. Even though some of these could have been retrieved from Google Maps, the validity of the toponymy was uncertain and so the original GLWD lake ID numbers were used to avoid confusion due to missing, multiple or uncertain lake names.

Insert Figure 2 near here

4.2 Selection criteria based on remote sensing
The criteria that satisfy the first objective of the selection protocol selected lakes conforming to (i) minimum detectable lake area filter, (ii) average number of daytime observations per year (‘pixel counts’), and (iii) water mask area compared to total lake surface area (fractional area, \( F \)). The automated facilities provided by ESA’s ATSR Reprocessing for Climate Lake Surface Water Temperature (ARC-Lake) project were used to apply these three remote sensing filters to ATSR2/AATSR archive data (cf. MacCallum and Merchant 2013). A description of these three selection criteria is presented below.

### 4.2.1 5 x 5 water cell filter

A minimum size spatial filter was applied to the image data to exclude water bodies that were too small or irregular for reliable detection and mapping with remote sensing mapping and thus having the highest likelihood of land contaminating pixels assumed to be water (resulting from geo-location inaccuracies). The filter consisted of a 5 x 5 pixel (equating to approximately 5 x 5 km) kernel and was developed as part of the ARC-Lake water detection scheme (MacCallum and Merchant 2013).

The ARC-Lake water detection scheme combines GLWD polygons and a binary land/water mask and was applied to the full ATSR2/AATSR time period (1995-2012), recording counts of positive water detection on a 1/120° grid for the entire period. The scheme was designed to overcome limitations arising from inaccuracies in GLWD polygons due to (a) mapping issues and (b) the representation of target areas for a single moment in time, which might not capture seasonal or long-term variability in surface area. Using the GLWD polygons as a basis for target location, but not limited by, a filter based on counts of positive water detection was applied on a target-by-target basis to reduce potential land contamination from mixed land/water pixels. Subsequently, a further image filter was applied to exclude targets where the largest individual area of water was smaller than a 5 x 5 area of contiguous water cells on the 1/120° grid.

### 4.2.2 Pixel count

To determine an estimate of the likely number of remotely sensed observations that could be expected from a lake in any one year (or period of years), results from the ARC-Lake water detection scheme were analysed across the complete ATSR2/AATSR archive (1995-2012). The automated water detection algorithm returned the number of daytime, clear-sky and ice-free observations of water (within lake) over the 18-year period, for each 1/120° grid cell (‘pixel counts’). This number is influenced by seasonal ice and cloud cover, and by changes in lake surface area. Larger values increase the likelihood that temporal coverage of image data in the archive is sufficient to estimate trends in satellite-retrieved physical parameters. Considering the maximum value of pixel counts across each potential target, a threshold of pixel counts ≥ 10 per year was used. This meant that only lakes with at least 10 daytime observations in at least one water cell per water body per year were retained on the candidate list. Due to cloud contamination, the maximum daytime observations per year in any one water body in this dataset was 42 (Fig. 3). The pixel count threshold was used to ensure that the selected lakes are likely to provide frequent enough observations (i.e. at least monthly, depending on cloud cover) to allow the study of seasonal patterns and lake phenology; an important aspect of long-term environmental change studies. However, whilst we have assumed a minimum of 10 observations per year, this value requires further critical
analysis, including whether key lake phenology events can be adequately captured, and is the subject of on-going work. Clearly though, for other applications that operate at different temporal scales the selection of this minimum number of observations will be critical in determining the viability and utility of a remote sensing-based approach.

Insert Figure 3 near here

4.2.3 Fractional area

Based on the pixel count data, a binary land/water mask representing the maximum water extent for each target over the entire 1995-2012 time period was generated for each water body on the list. Using this information, the ratio, $F$, of the water mask area to the total surface area of the water bodies under investigation was calculated (called ‘fractional area’ from hereon) to investigate what percentage of the lake surface area could repeatedly be remotely sensed using daytime satellite (ATSR2/AATSR) observations. A threshold of 30% was set as a minimum ratio for this criterion, meaning that at least 1/3 of the lake should be remotely sensed over the 1995-2012 period. This threshold ensures that a large enough area of the lake is monitored through the years, accounting for bias that can arise from extracting data from only small parts of a lake, particularly when sub-basins with potentially distinct ecological characters exist in a single lake.

This three-step filtering process identified a subset of the GLWD Level-1 inventory, which we have called ‘Preliminary Sample’ from here onwards.

4.3 Selection criteria based on lake landscape context

In order to fulfil the second objective of this work, which was to maximise the sample of lake types and behaviours from across the global continuum, a hierarchy of geographical, landscape setting and lake morphology attributes were considered (cf. WFD 2000; Soranno et al. 2009). Specifically, lakes were selected based upon two criteria: (i) water body type and (ii) lake-catchment relationships, whilst (iii) shoreline irregularity, (iv) eco-location, and (v) basin morphometry classes were used to test the degree of representativeness of each class. Seasonality and lake water quality were also considered as selection criteria, but this was hampered by a complete lack of this type of information at a global standardised scale. Nevertheless, the process was designed to include lakes with highly variable surface area (e.g. Aral Sea) and no restrictions were placed on surface area variability, resulting in the inclusion of targets with large seasonal or long-term variations in surface area (e.g. Lake Balqash, Kazakhstan and the Great Salt Lake, USA; see Supplementary Material for entire list of ephemeral sites). What is more, the variable landscape setting of the candidate water bodies acts as an indication of expected variability in the ecological characteristics of these sites, and particularly their pH, alkalinity, eutrophication status and mixing regime.

4.3.1 Water Body Type

The GloboLakes project will focus on inland lentic waters that have no (or very limited) connection to and interaction with the ocean. As a result, natural fresh and saline lakes, reservoirs and coastal lagoons were included, whilst rivers, estuaries, fjords, glaciers and coastal bays were excluded. All water bodies that appeared to be dry or semi-dry, or had a basin that could not be visually distinguished on Google™ Earth, were flagged as
ephemeral. An example is shown in Fig. 4. Potentially ephemeral lakes were also identified by their location in arid areas of the world. Finally, all coastal lagoons in the preliminary sample were identified and classified according to Kjerfve’s (1986) geomorphic classification. Only choked lagoons or those unconnected to the ocean, which have water body characteristics dominated by terrestrial inflows and mixing/stratification behaviours consistent with inland freshwater lakes were identified and removed. A total of 132 lagoons were identified, 22 of which were discarded as belonging to the ‘restricted’ or ‘leaky’ type. The remaining 110 ‘unconnected’ or ‘choked’ lagoons were retained for further consideration.

4.3.2 Lake-Catchment Relationships

The relationship of a lake to its catchment is a key factor influencing to a lesser or greater extent the physical, chemical and biological attributes of the water body due to its natural role in runoff, sediment and nutrient supply and susceptibility to human-induced pressures (including indirect effects of climate change) arising from land use change, water resource management and biodiversity impacts (Schindler 2009). The GloboLakes study site selection therefore incorporated lakes with a wide range of catchment-to-lake-surface-area ratios.

However, in the case of lagoons, only large catchment areas were selected to ensure that the surrounding land is responsible for the majority of the ecological processes affecting the lagoon and that the effect of the adjacent coastal waters is limited. This only applied to lagoons that are connected to the sea, so for this step all unconnected lagoons remained on the list despite the size of their catchment. A threshold of ten was used for the catchment-to-lagoon-area ratio ($R$) resulting in the exclusion of 65% of the preliminary sample choked lagoons that are connected (whether permanently or not) to the sea.

By applying these two criteria to the data, the preliminary sample was further reduced to 961 water bodies. The Caspian Sea, with a surface area equal of 378,119 km$^2$, was removed as an excessively large outlier, which is generally treated as sea in most global remote sensing and modelling systems. The remaining 960 water bodies (which include 805 lakes, 122 reservoirs and 33 lagoons, including 41 ephemeral sites as identified in this study; Fig. 5(a)) will be referred to ‘GloboLakes sites’ from here onwards; a complete list of these lakes with their names, coordinates, country and water body type is provided in the supplementary material.

4.4. Representativeness of selected lakes

The resulting GloboLakes sites were analysed using the three criteria described below to assess how well each criterion-relevant class is represented in the sample.
4.4.1 Shoreline irregularity

The Shore Development Index (SDI) was used as a measure of shoreline irregularity and is the ratio of the shore length (perimeter of lake), $L$, to the length of the circumference of a circle of area equal to that of the lake, $A$, (Hutchinson 1957):

$$SDI = \frac{L}{(2\pi A)^{0.5}}$$  \hspace{1cm} \text{Equation (1)}

The SDI is dimensionless and generally ranges between 1 (perfectly circular lakes) and 10 (highly irregular lakes), with values over 10 being less common (Håkanson 2004). As there is no known published classification of the regularity of lakes according to their shoreline development index, in this study we used the classes shown below:

- Circles; SDI = 1
- Circular; $1 < SDI \leq 3$
- Regular; $3 < SDI \leq 5$
- Irregular; $5 < SDI \leq 8$
- Highly irregular; $SDI > 8$

The SDI scores for all GloboLakes sites were calculated using equation (1). Most GloboLakes sites (91.1%) have circular to regular shorelines, an optimum shape from a remote sensing perspective, whilst only 84 sites are irregular or highly irregular.

Investigation of the data showed no correlation between high SDI scores and low pixel count values (linear $R^2 = 0.0021$). Lower pixel counts were mostly observed at high latitudes, where the periods of lake ice cover last for longer, whilst the SDI depends on the lake origin and landscape, with high SDI scores most common amongst reservoirs.

As a result, the pixel count filter used in this study did not seem to impose bias on the SDI distributions of the GloboLakes sites.

4.4.2 Eco-location

Ideally, the GloboLakes study lakes should seek to represent the widest possible geographical distribution, with lakes included from all biomes and climatic zones in the world. The ecoregion of each lake, along with topographic information (altitude) and geographical location (latitude) was used to classify lakes according to their ‘eco-location’. The Terrestrial Ecoregions of the World (TEOW) database (Olson et al. 2001) was used to identify different terrestrial ecoregions and ensure all (or the majority) are represented. With the exception of ‘Rock and Ice’, all TEOWs are represented by the GloboLakes sites (Fig. 6). In the GLWD Level-1 database, only three lakes are situated within this ecoregion, two of which were identified as glaciers in this work (and consequently discarded) and the third was filtered out during the remote sensing-based selection process. ‘Boreal, Taiga’ and ‘Tundra’ are by comparison under-represented in the GloboLakes sites due to these lakes being subject to long freezing periods and frequent cloud cover (and subsequently low pixel counts). By contrast, montane and temperate grasslands seem to be generously represented in the GloboLakes sites. This is probably because most GLWD Level-1 water bodies located within ‘montane grasslands’ (97%) and ‘temperate grasslands’ (90%) are lakes with circular to regular shorelines ($1 < SDI \leq 5$) and thus easily detectable with remote sensing, which increased their chances of being selected in this study.

Between the Southern and Northern Hemispheres, the latitudinal distribution of the GloboLakes sites is asymmetrical, with the vast majority (> 85%) found in the Northern Hemisphere and most of them above 30°N. There are no GLWD Level-1 lakes (and in
extension GloboLakes sites) situated below 60°S (Fig. 6), i.e. Antarctica. Compared with GLWD Level-1 inventory, there seem to be relatively few high northern latitude lakes (60-90°N) in the GloboLakes sites due to issues of frequent cloud and long ice cover, which contrasts with the presence of relatively more water bodies in the latitude band 30-60°N, when compared to the GLWD Level-1 distribution.

According to the EU Water Framework Directive (WFD) typology for altitude, lowland lakes lie below 200 m, mid-altitude lakes lie between 200-800 m and high-altitude lakes lie above 800 m (WFD 2000). Almost half of the GloboLakes sites are lowland (45%), one third (32%) are mid-altitude and the rest (23%) are high-altitude water bodies (Fig. 5(b)). Of the 432 lowland water bodies, 30 are situated below 0 m and 34 at 0 m. Compared with the GLWD Level-1 inventory, the GloboLakes sites contain relatively fewer mid-altitude and more high-altitude lakes. Investigation showed that 57% of the mid-altitude lakes that were discarded during the selection process, are smaller than 100 km², which limits their detectability with remote sensing, whilst 12% of the larger (≥ 100 km²) ones are situated at high latitudes (north of 60°N), which reduces their average pixel count per year due to long freezing periods and frequent cloud cover. On the other hand, high-altitude lakes generally exhibit circular to regular shorelines (94% of total high-altitude GLWD Level-1 water bodies), which increases their detectability with remote sensing.

Insert Figure 6 near here (preferably occupying a full page in landscape mode)

4.4.3 Basin morphometry

The candidate list of lakes should ideally include lakes with variable morphological characteristics, incorporating relatively small to medium-sized lakes and some of the largest lakes in the world in terms of surface area and volume, as well as shallow and deep lakes. The GLWD Level-1 inventory lists the largest 3,721 lakes and reservoirs in the world, including lakes with surface area greater than 50 km² and reservoirs with storage capacity greater than 0.5 km³. According to the Annex A of the European Union (EU) Water Framework Directive (WFD 2000) lake typology for surface area, all GLWD Level-1 water bodies are classified as large (> 0.5 km³). The GloboLakes sites contain relatively few lakes with surface area less than 100 km² and more towards the middle and upper end of the size classes. The latter reflects the limitations of currently available remote sensing systems with respect to monitoring (relatively) small lakes, especially when their shape is irregular.

According to the EU WFD typology for depth, very shallow lakes have a mean depth of less than 3 m, shallow lakes between 3-15 m and deep lakes have a mean depth greater than 15 m (WFD 2000). There are no data for mean depth in the GLWD database, but for 151 water bodies the information was derived from an extensive literature research and from existing online databases. Fig. 5(c) shows that of the GloboLakes sites with known mean depths, most are either classified as shallow (45%) or deep (41%). Similarly, information on volume was unavailable for most of the GLWD Level-1 water bodies (82%) and GloboLakes sites (78%). However, the 205 lakes for which volume data were available, or were found in the literature or online, cover the entire continuum with the majority (68%) being between 2-100 km³ (Fig. 5(d)).
5. Discussion and Conclusions

The aim of this work was to use a combination of remote sensing techniques and lake typology criteria in order to select a list of globally distributed lakes and reservoirs that (a) are appropriate for an environmental change study based on remote sensing data and (b) span a wide range of lake characteristics. The long archive of ATSR2 and AATSR sensors (1995-2012) on-board two European Space Agency satellites and a comprehensive lake database that lists the largest lakes and reservoirs in the world were utilised for this purpose. The methodology proposed here combined spatial filters of remotely sensed detectable area of lake water accounting for land and cloud contamination, and number of daytime satellite observations per year. In addition, auxiliary information sourced in GLWD and other online databases helped inform the site selection with the use of a number of typology criteria, including lake and catchment morphological data and ecological setting. A total of 960 lakes, lagoons and reservoirs were selected using this process with surface areas between 48-82000 km$^2$ and spanning a wide range of ecological characteristics fulfilling the two main aims of this study.

Uncertainty related to the GLWD data meant that some of the available information was used with caution in this work. For example, the GLWD polygons only represent the target area for a single moment in time, and therefore do not capture seasonal or long-term changes in surface area and may not even provide an accurate representation of the lake area for any time in the ATSR2/AATSR observing period. As a result of this, the development of a water detection algorithm for the ATSR2/AATSR archive (MacCallum and Merchant 2012), which could then be used for the site selection, was an essential process for this work. In addition, the GLWD catchment area and altitude estimations were based on a rather coarse dataset at 1 km$^2$ spatial resolution, and their associated uncertainty is due to scaling issues and model inaccuracies (Lehner and Döll 2004). Ongoing work within the GloboLakes project aims to address all these issues.

The relative paucity of global standardised lake and catchment data that are currently available was one of the main limitations of this work. The unavailability of lake depth, volume, seasonality and water quality data for most lakes across the globe, and the unknown reliability of some of the existing data, restricted the application of all selection criteria for the second objective, which was to select water bodies that span a wide range of ecological behaviours. Based on global datasets, only the eco-location of study sites could be derived with confidence. The GloboLakes project aims to deliver most of this missing information, including modelled lake morphometry (mean, maximum depth and volume) and lake water quality data.

Even though SDI is a useful measure of lake shoreline irregularity, and can be used as an estimation of effective area of open water available for remotely sensed observations, it should be used with caution. Some of the lakes that were classified as ‘highly irregular’ are the largest lakes (by surface area) of the world, such as the Great Slave Lake, Lake Chad and Lake Nettling. These three lakes show localised irregularity that results in high SDI scores, but have large enough basins that permit reliable and repeatable remotely sensed observations of a considerable proportion of the lake surface. To overcome problems like these when selecting candidate lakes, the pixel
count filtering technique is an essential tool to estimate the effective area of open water available for remotely sensed observations.

The final sample of GloboLakes sites has a high degree of variability in their landscape context as defined by ecoregion setting, landscape position, lake-catchment relation and, finally, lake morphology. These systems embrace a wide range of formative mechanisms (e.g. volcanic, glacial, tectonic, fluvial, etc.) and feature a spectrum of human pressures – from essentially wilderness settings and natural conditions to intensely developed, highly regulated and modified conditions. Accordingly, they provide a unique testing opportunity to evaluate trends in water quantity, quality and ecosystem response as a result of changing pressures, and as a direct and indirect consequence of climate change.

Whilst currently constrained by the need to work on systems of a particular size and shape that lend themselves to analysis using the current generation of Earth Observation (EO) tools, the future of global lake analysis with remote sensing looks much brighter as the next generation of platforms (e.g. the Sentinel series of satellite sensors) are launched and commissioned. The new Sentinel satellites will ensure spatially, temporally and spectrally enhanced data continuity as they build on the success of current and past sensors. For example, the SAR system on Sentinel-1 and the SPOT- and Landsat-like data from Sentinel-2 will make it possible to track variability in lake area. However, despite improved spatial, spectral and radiometric resolutions of new missions, the issues highlighted in this work will be important considerations for the application of systematic retrieval of lake water quality from satellite-based observations. In particular, we have shown that lake size, shape, cloudiness and the availability (or lack of it) of lake contextual information, can inhibit the utility of satellite-based observations, and such considerations will remain for future missions.

The methodology proposed here should be of value to all researchers hoping to exploit archive and future satellite-based missions for lake studies. In particular, it will be applicable to the Sea and Land Surface Temperature Radiometer (SLSTR), which will be the successor to the (A)ATSR, while the size and shape of the lakes selected will be appropriate for use with OLCI data; both of which are carried on-board Sentinel-3 that was launched in February 2016. Based on these improved satellite sensor technologies, future application of the proposed site selection protocol will enable the selection of much smaller lakes than possible before. In addition, the use of such datasets will be further enhanced with new data handling and sharing protocols providing unprecedented interoperability. In particular, the ESA Copernicus Programme (http://www.copernicus.eu/) provides free data access and the EU INSPIRE directive (http://inspire.ec.europa.eu/) aims to improve the availability, quality, accessibility and sharing of data across Europe. All of this points to a significant potential of remote sensing for regional to global limnological analysis but key questions still remain to be resolved. The most important of these is that of whether sufficient observations exist to allow retrieval of lake phenology. We have assumed that a minimum of 10 observations per year will allow this, but this requires further analysis which we are currently undertaking, and will be a key output of the GloboLakes project.

Acknowledgements
This work was supported by the United Kingdom Natural Environment Research Council (NERC) under Grant NE/J024279/1. The authors would like to thank University of Edinburgh and the ARC-Lake project for access to data, and acknowledge our partners in the GloboLakes Consortium, namely Universities of Stirling, Reading and Glasgow, Plymouth Marine Laboratory (PML) and the Centre for Ecology and Hydrology (CEH).

References


ESA (European Space Agency) ArcLake (ATSR Reprocessing for Climate: Lake Surface Water Temperature & Ice Cover) project 2016. Accessed online February 16, 2016 at http://www.geos.ed.ac.uk/arclake


Balaton using 10 years of MERIS observations” Remote Sensing of Environment 158: 441-452.


USFS (US Forest Service) FIA (Forest Inventory and Analysis) programme 2016. Accessed online February 18, 2016 at http://www.fia.fs.fed.us/


List of figure captions

**Fig. 1:** Flowchart presenting the site selection process that was developed in this study.

**Fig. 2:** Map showing the location of 51 lake outlines identified as ‘circles’ on GLWD Level-1 inventory, including zoom-in windows in three locations.

**Fig. 3:** Frequency of pixel counts per lake per year.

**Fig. 4:** Examples of cases that were flagged as ephemeral, because they appeared to be dry or semi-dry (Lake ID408), or had a basin that could not be visually distinguished on Google™ Earth (Lakes ID799 and ID3575).

**Fig. 5:** Pie charts showing relative percentages (%) of GloboLakes sites based on: (a) water body type, (b) lake elevation, (c) lake mean depth (where available), and (d) lake volume (where available).

**Fig. 6:** Map showing the spatial distribution of the GloboLakes sites globally, based on the Terrestrial Ecoregions of the World (TEOW) (Olson et al. 2001), lake surface area (variable circle diameter) and Shoreline Development Index (SDI) regularity class (in different shades of blue).
Fig. 1: Flowchart presenting the site selection process that was developed in this study.  
182x282mm (300 x 300 DPI)
Fig. 2: Map showing the location of 51 lake outlines identified as ‘circles’ on GLWD Level-1 inventory, including zoom-in windows in three locations.

210x148mm (300 x 300 DPI)
Fig. 3: Frequency of pixel counts per lake per year.
101x81mm (300 x 300 DPI)
Fig. 4: Examples of cases that were flagged as ephemeral, because they appeared to be dry or semi-dry (Lake ID408), or had a basin that could not be visually distinguished on Google Earth (Lakes ID799 and ID3575).

210x148mm (300 x 300 DPI)
Fig. 5: Pie charts showing relative percentages (%) of GloboLakes sites based on: (a) water body type, (b) lake elevation, (c) lake mean depth (where available), and (d) lake volume (where available).

209x297mm (300 x 300 DPI)
Fig. 6: Map showing the spatial distribution of the GloboLakes sites globally, based on the Terrestrial Ecoregions of the World (TEOW) (Olson et al. 2001), lake surface area (variable circle diameter) and Shoreline Development Index (SDI) regularity class (in different shades of blue).

296x210mm (300 x 300 DPI)