A Tool Supporting the Extraction of Angling Effort Data from Remote Camera Images

Saul Greenberg
Department of Computer Science, University of Calgary,
2500 University Dr NW, Calgary, Alberta, Canada, T2N 1N4. +1 403 220 6087
Email: saul.greenberg@ucalgary.ca

Theresa Godin
Freshwater Fisheries Society of BC, Research Evaluation and Development Section,
Room 342, 2202 Main Mall, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4
Email: theresa.godin@gofishbc.com

Cite as:

Greenberg, S. and Godin, T. *A Tool Supporting the Extraction of Angling Effort Data from Remote Camera Images.* Technical Report 2013-1050-17, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada. T2N 1N4.

Abstract

Angling effort is often estimated from data captured by creel survey (prohibitively expensive to do on more than a few lakes), or aerial surveys (limited to summer effort estimates). A recent alternative method uses remote cameras to capture images of lakes at hourly intervals over long time periods. Technicians then visually analyze the thousands of generated images for features of interest (e.g., angler counts and environmental conditions), and use that data to estimate angling effort. The problem is that the visual analysis step is time-consuming, expensive, error-prone, and difficult to validate. Consequently, we elicited the strategies and best practices technicians used when analyzing images, and identified bottlenecks. We then designed software – called TIMELAPSE – to better support image analysis. In use for several years, TIMELAPSE has proven cost-effective: it significantly eases a technician's workflow while reducing errors. TIMELAPSE is now an effective part of estimating angling effort in BC's small lakes fisheries.

Introduction

Freshwater fishing effort is one of the important metrics used by fisheries biologists to manage recreational fishing on small lakes in British Columbia. One of two strategies, or both, is typically employed to determine effort depending on the objective. First, *aerial boat counts* provide a standard index of angler activity for many lakes across a broad landscape over years (Tredger 1992, Parkinson et al. 1988) and is a useful component of a larger sampling program. While the cost per lake is reasonable, total effort estimates for individual lakes are highly variable, and flight counts for this program are limited to mid-day, weekend, and fair weather summer days in non-mountainous terrain. Second, stratified *creel surveys* provide detailed information on specific fisheries, but their cost limits the number of lakes that can be monitored.

A third strategy has been developed to estimate angling effort: time-lapse *remote field cameras* that capture hourly images of lakes over long time periods. While remote cameras have broad scientific use in wildlife research (Kays and Slauson, 2008; O'Connell, Nichols and Karanth, 2011), they have only recently been applied to fisheries. Technicians visually analyze each image to extract relevant data, which in turn is used to calculate angling effort (Newton et. al, 2013; van Poorten et. al., in preparation). This approach has many advantages. Cameras are suitable for monitoring remote and/or hard-to-access lakes (found in the interior and northern half of BC), for capturing winter effort (important as ice fishing can represent a significant proportion of resource pressure), and for capturing daily and seasonal effort trends. As well, camera data can be used alone, or combined with data collected from the other two strategies to produce an even more accurate view of angling effort over time.

Using cameras to calculate fishing effort comprises four primary steps.

1. **Camera placement and image capture.** Cameras must be deployed to capture a reasonably broad field of view of the lake, while still producing images of sufficient fidelity for an analyst to visually detect and discriminate between objects of interest (Newton et. al 2013). Placement must also consider public concerns about privacy, where cameras should avoid capturing high fidelity features of people. In practice, these constraints favor camera positions that capture relatively broad fields of view of small lakes (see Figures 1 – 3). Technicians affix cameras to landscape features (typically trees) at one or more strategic locations surrounding a lake. Each camera is configured to repeatedly capture images over time − usually once every hour − for periods of weeks or months. Alternately, if placed at a lake's access point, image capture can be configured to use motion detection.

2. **Image retrieval and storage.** Technicians revisit cameras periodically, usually every 4 to 10 weeks, to change camera batteries if necessary, and to retrieve images. A single camera's images are stored within a labelled folder – an *image set* – whose unique name comprises the specific lake and date range.

3. **Image analysis.** An analyst – a fisheries technician, biologist, or contractor – visually scrutinizes every digital image in an image set for particular features of interest. For each image, she encodes information such as the lake identification, time and date the image was captured, camera name, environmental conditions, and angler / boat counts. Particular features of interest vary with the study objective. Completed spreadsheet data is eventually read into a database.

4. **Calculating recreational angling effort.** The fisheries biologist calculates an estimate of angling effort from this raw data, and can calibrate it against existing creel and aerial survey results if available (Newton et. al 2013, van Poorten, et al). Doing at least 10 non-zero instantaneous angler counts concurrently at times when the images are taken improves the estimate. Broadly speaking, this provides a ratio or correction factor of the number of anglers seen by the camera to the angler numbers on the total lake at one time. If instantaneous counts are unavailable, image counts are still useful for comparing relative effort across seasons.

A bottleneck in this process – and the focus of this article – is in the third image analysis step. Each image set comprises a large number of images − typically hundreds but usually thousands − and there are a multitude of image sets. For example, BC currently monitors from 30 to more than 70 small lakes throughout the province using one (and sometimes two) cameras per lake set to take hourly images. This generates roughly several hundred thousand to more than half a million images per year, each which must be systematically stored and hand-analyzed. The problem is that image inspection and encoding takes considerable time, is tedious and error-prone, and is hard to validate. The costs for analyzing images alone makes effort estimates obtained using cameras reasonably expensive.

Consequently, the Freshwaters Fisheries Society of BC decided to investigate and improve upon this important step. Our first goal was to understand the existing *ad-hoc* image analysis process. We used a contextual inquiry method (Holzblatt et. al., 2004), where we interviewed and stepped through the existing process with various analysts to uncover their strategies and to detect pinch-points. Our second goal was to develop a software tool to support the analyst's best practices, ideally resulting in an efficient image inspection and data encoding practice. As we will see, the same system can be used to verify the reliability of entered data, could also serve as a platform for citizen science for crowd-sourcing image analysis, and is potentially applicable to a broad range of other wildlife and fisheries-related projects involving remote cameras.

Study Sites and Typical Images Captured

The 70+ lakes currently monitored in BC differ considerably in their characteristics as well as their angling effort. The several examples below illustrate their variance. Lake characteristics constrain camera placements, the type of image produced, and differences in fidelity of objects of interest. They also illustrate why image analysis can be difficult.

High effort urban lake: camera capturing activities around dock.

Close to a dozen BC lakes are managed as urban lakes, i.e., a lake situated in a high-density urban area that receives considerable visitation. Urban lakes tend to be small (<50 hectares). Most are heavily managed for multiple recreational activities (biking, dog walking, picnicking, and fishing). They often have good access points (parking, walkways), well-developed pathways around the shoreline (allowing fishing access from various points on the shore), and docks that further concentrate boating and fishing activities. Figure 1 illustrates a remote camera on a typical urban lake, positioned primarily to capture the concentrated fishing activity on the dock, but also fishing from boats and from the far shoreline.

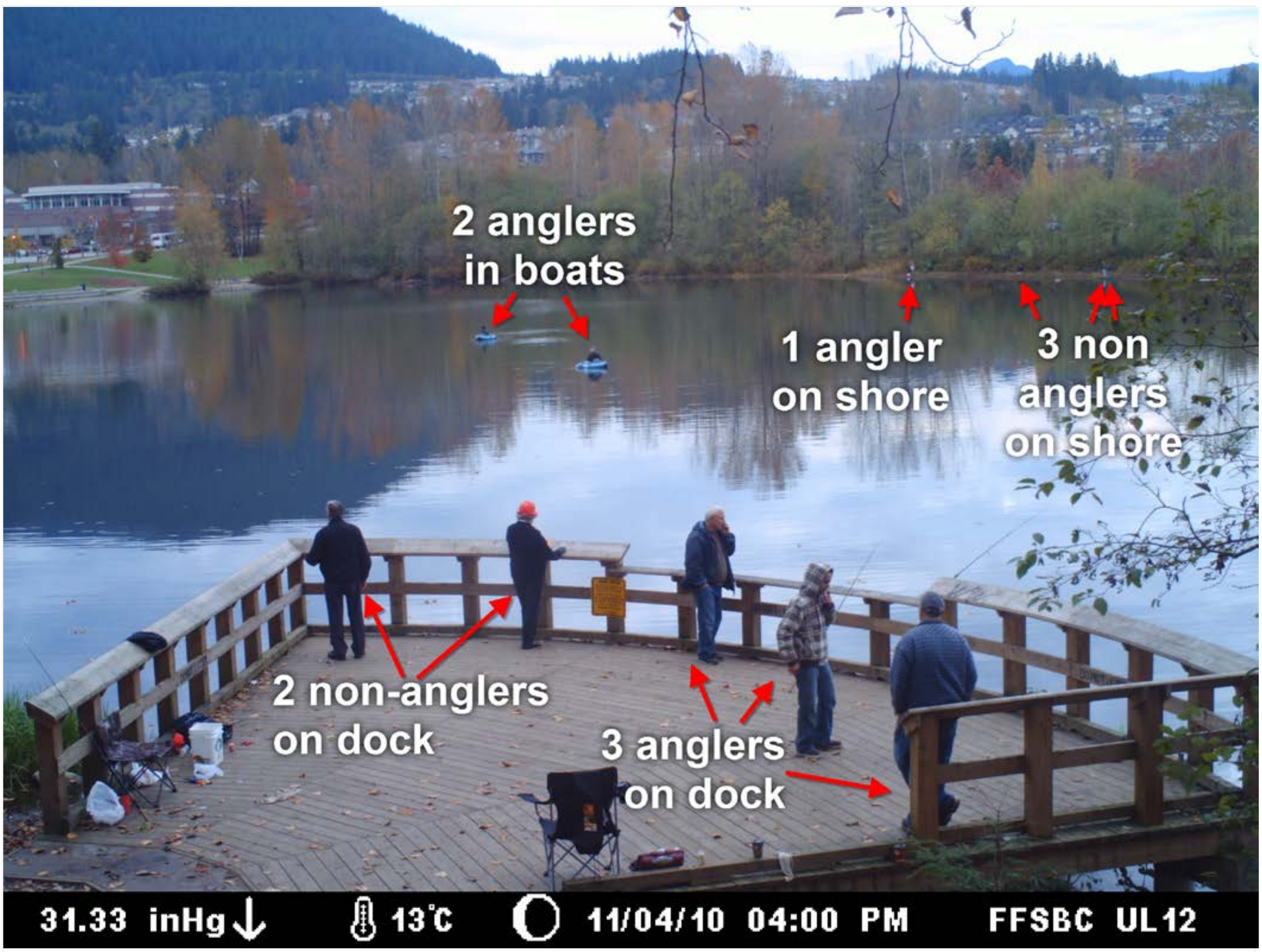

Figure 1: High effort urban lake, with camera capturing activities primarily around dock.

Low to moderate effort remote lake: camera captures majority of lake surface and accessible shoreline.

The majority of BC lakes are in 'remote' rural or wilderness areas. Access ranges from paved to 4x4 roads, to walk-ins. While angling effort typically ranges from low to moderate, its determination is still critical, especially if vulnerable species are involved. Figure 2 illustrates a camera positioned on a typical remote lake to capture ~50% of the surface area (for capturing boats) and the majority of accessible shoreline (for capturing shore anglers) including its campsite, boat launch, and primary access point.

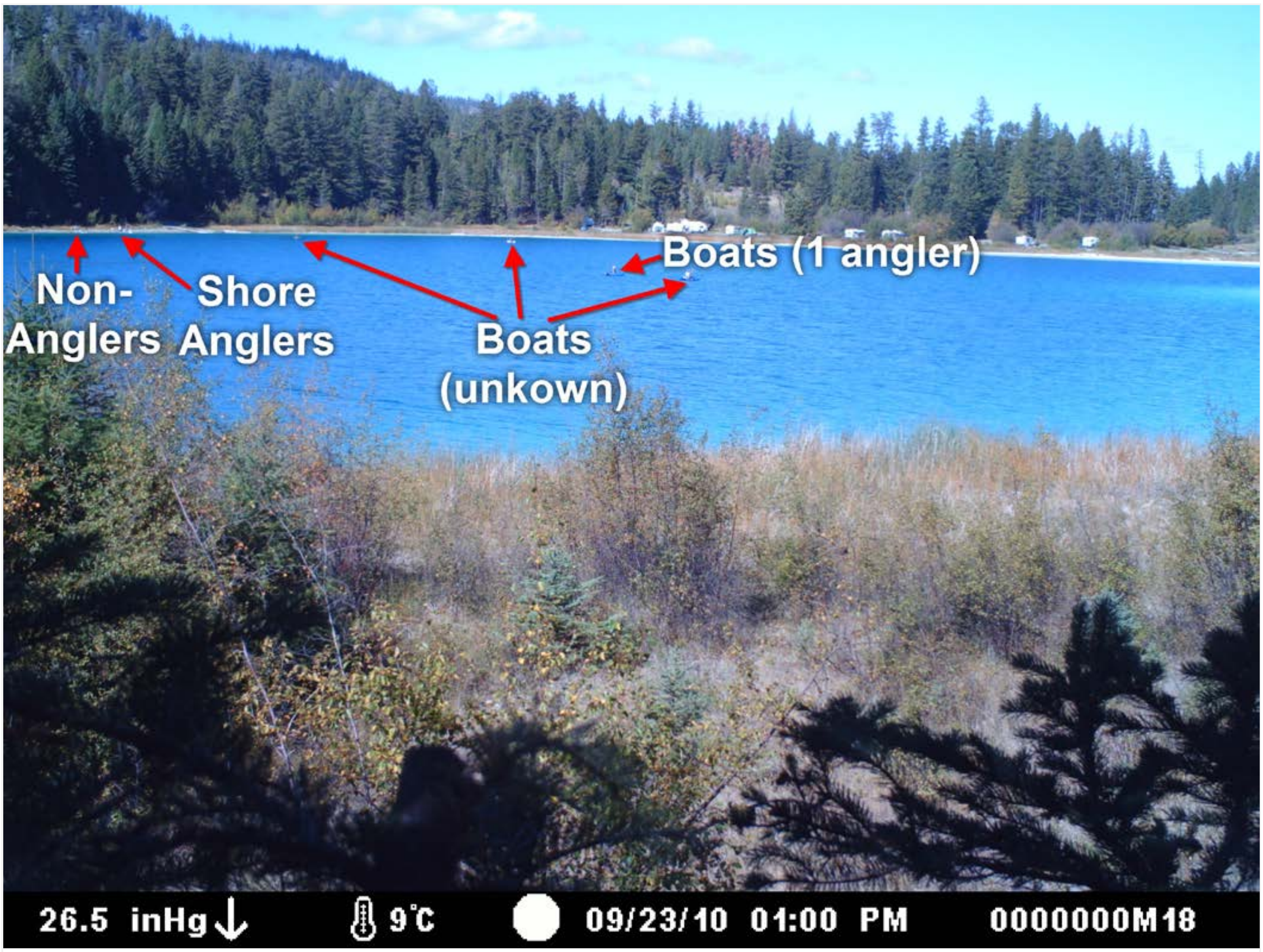

Figure 2: Moderate effort remote lake, with camera capturing roughly 50% of lake surface and accessible shoreline.

A high effort winter fishery, with camera capturing commonly fished area of ice surface

In some lakes in northern regions, the 4-month winter fishery represents a significant proportion of a lake's total effort. Yet winter conditions means that aerial surveys are not used and creel surveys only occasionally done. Figure 3 illustrates an image from a winter fishery camera, where it captured the most commonly fished area on the ice surface, including ice anglers, their companions and pets, and the equipment they bring to ease transport and fishing comfort. Ice fishing is often concentrated in hotspots so that a camera capturing 30% of lake surface can capture >50% of the angling pressure.

Other Study Site Factors

Not all study sites precisely fit the above descriptions. For example, for larger lakes or lakes with multiple basins, one or two cameras cannot capture a large enough proportion of anglers to estimate total effort. Instead, cameras – configured in motion detection mode – can be placed at access points including boat launches and trails to estimate arriving and departing anglers or fishing hot spots. As another example, cameras can be positioned to best capture images that reveal other data on resource use, such as demographics (gender, age group) and activities (e.g., biking, hiking, and swimming).

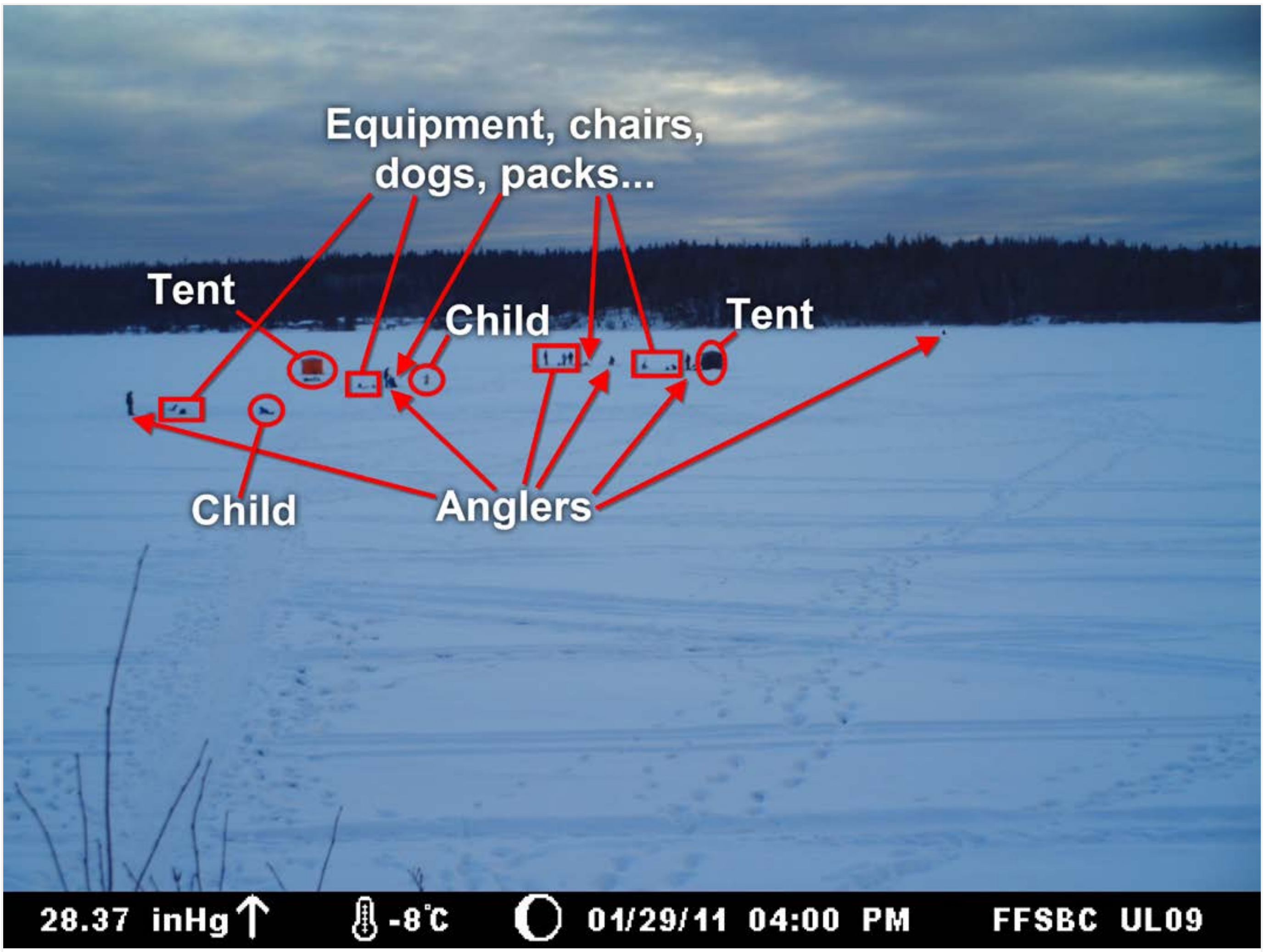

Figure 3. A high effort winter fishery, with camera capturing the most commonly fished area on the ice surface. This image captures a mix of anglers at different locations and distances, fishing paraphernalia, pets, and children.

Mitigating Problems in the Image Analysis Process

We walked through the existing image analysis process with various analysts to uncover the strategies they were using, how they accomplished those strategies, and to detect pinch-points. At a high level, image sets were distributed to analysts along with a Microsoft Excel spreadsheet template: each spreadsheet row represents a single image, while its columns indicate data categories. Using off-the-shelf photo software such as Microsoft PhotoViewer, analysts then inspected each image for various attributes, and recorded any results in the corresponding spreadsheet row (e.g., date, time, anglers, boats, environmental conditions…). After completion, the analyst returned the spreadsheet to the biologist. The biologists then inspected that data for possible problems, and if none were detected, used that data to calculate effort. Our walkthrough identified significant problems and inefficiencies: the process proved tedious, time-consuming and effortful. Data was fraught with accuracy issues and errors, and was difficult to validate. Consequently, we designed TIMELAPSE, an image analysis software tool that mitigates various workflow problems. TIMELAPSE and documentation on how to use it are available as free downloads (Greenberg, 2013; Greenberg and Godin, 2011).

The paragraphs below roughly follow the image analysis process, identifying particular problems and how they are mitigated by TIMELAPSE.

Data collection requirements.

The biologist in charge initially decides what data should be collected from the image sets, and communicates those needs to the analysts. While some data requirements are ubiquitous across projects, others are specialized to particular projects. The existing process specified data requirements as column headers in a spreadsheet that analysts would then fill in. As such, the meaning of these headers, the data type, and the data format required was sometimes difficult for the analyst to understand or remember. Analysts had to be especially vigilant in entering data into the correct row and column. These led to a variety of errors, which typically required the image set to be corrected or even re-analyzed.

To mitigate these problems, we developed a *data template*: a standardized, well-formed data description file where the biologist specifies exactly what data is required for each image, how data fields should appear in the user interface (e.g., their labels and tooltip help contents), and how the output data table should be constructed. They do this using the Timelapse Data Creator Tool (Figure 4), where each desired data element is created as a row in a table. The first five data elements are mandatory, as they specify data that biologists typically require across all projects: the *date* and *time* the image was taken, the image's *file name* and *folder name*, as well as the *image quality*. The remaining items are custom-created by the project's biologists, where each item conforms to one of three generic 'types' (see Figure 4, 2^{nd} column). *Counters* contains counts of particular features identified in an image, *notes* contain free-form text entries, and *fixed choices* are constrained to values selected from a given list (a menu) of values. Other fields specify how that data appears and behaves in the user interface, and how data output is labeled.

The biologist then copies the data template file into each image set's folder, where TIMELAPSE reads in the data template to create a project-specific interface. Figure 5 illustrates this, where the data fields at its top were constructed from the data template in Figure 4.

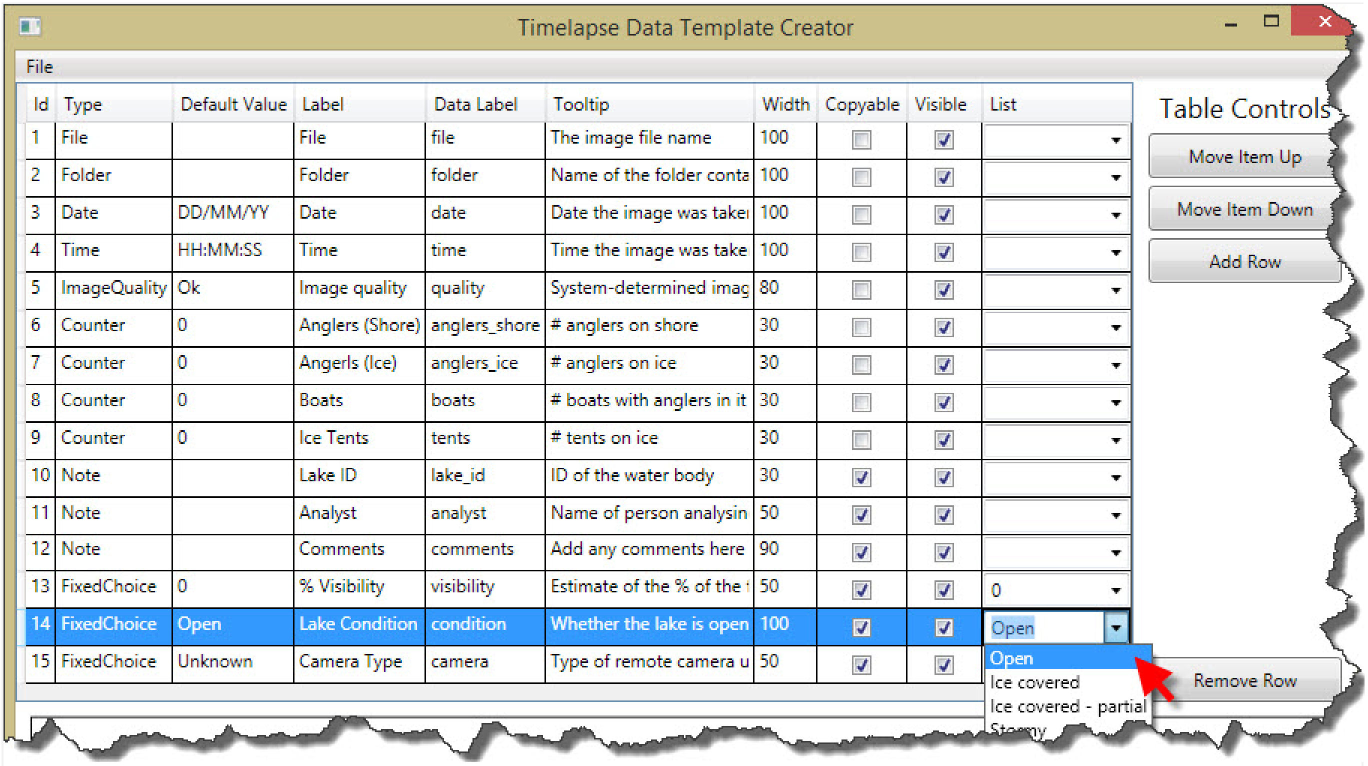

Figure 4. Data Template tool, where the biologists creates all the data fields of interest along with attributes that indicate how it will appear in the Timelapse user interface. This illustration shows a simpler subset of the data fields actually used by the authors.

In our own practice at the Freshwaters Fisheries Society of BC, the data template proved invaluable as a way to structure data, where the template became the input standard applied to all image sets captured across many lakes, and the output standard to facilitate uploading data to a provincial small lakes database.

The first pass: entering routine housekeeping image data.

When analysts receive an image set, they manually enter routine housekeeping data describing each image: its file name and the name of its containing folder, the time and date that image was taken, and the image quality. This step is particularly time-consuming and tedious.

TIMELAPSE eliminates this manual chore from the analyst's work flow by automating this first step, thus saving considerable time. When TIMELAPSE initially opens an image set, it extracts information from each image's meta-data and uses that to populate its file and folder name, and its time and date. Using image analysis, it also checks the image quality to see if it is a nighttime shot or otherwise corrupt. This data is displayed in the TIMELAPSE interface on the first data row (Figure 5). If problems are found (e.g., because of ambiguities in how different cameras specify dates), TIMELAPSE will query the analyst with ways of resolving it. We note that other information – depending on the camera – is also available in the metadata (e.g., temperature, barometric pressure) and could be extracted if needed.

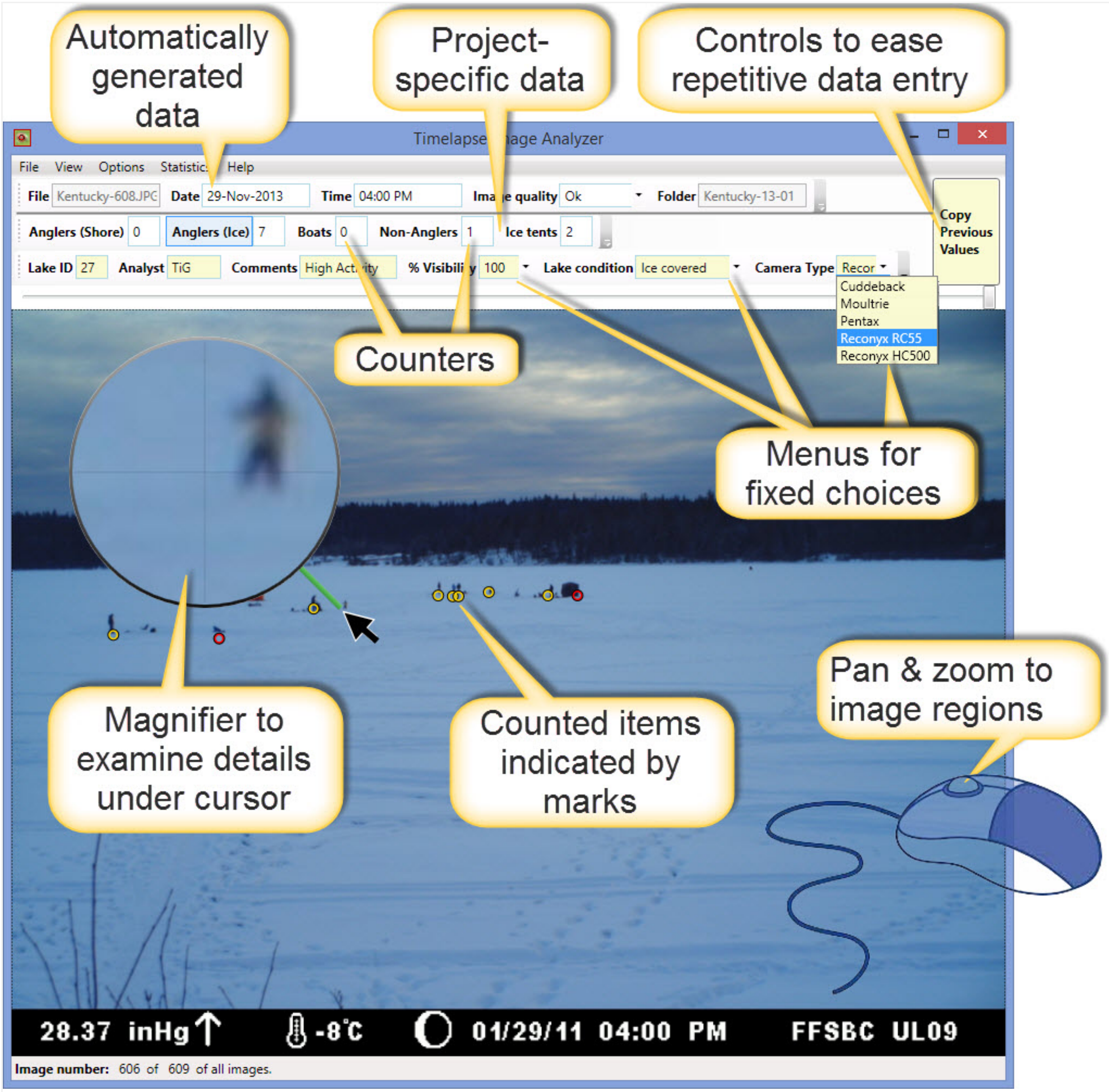

Figure 5. An annotated screenshot of the Timelapse in action, displaying the image in Figure 3.

Visually searching and inspecting each image.

An analyst's primary task is to visually inspect each image for important features of interest. While some features are determined by examining the image as a whole (e.g., lake conditions), others require searching the image to find features of interest. This can be difficult with off-the-shelf photo viewers. When a camera captures a broad field of view, features of interest (such as a distant on-shore or ice angler, or a distant boat) may be quite small and difficult to spot, as evident in Figures 1 to 3. This is especially true if activities on that lake are rare, and if lighting, fog, or shadow compromises image quality. Even when a feature is seen, the analyst must inspect it to classify what it is (e.g., an angler *vs*. a non-fishing person, *vs*. a chair *vs*. a shadow).

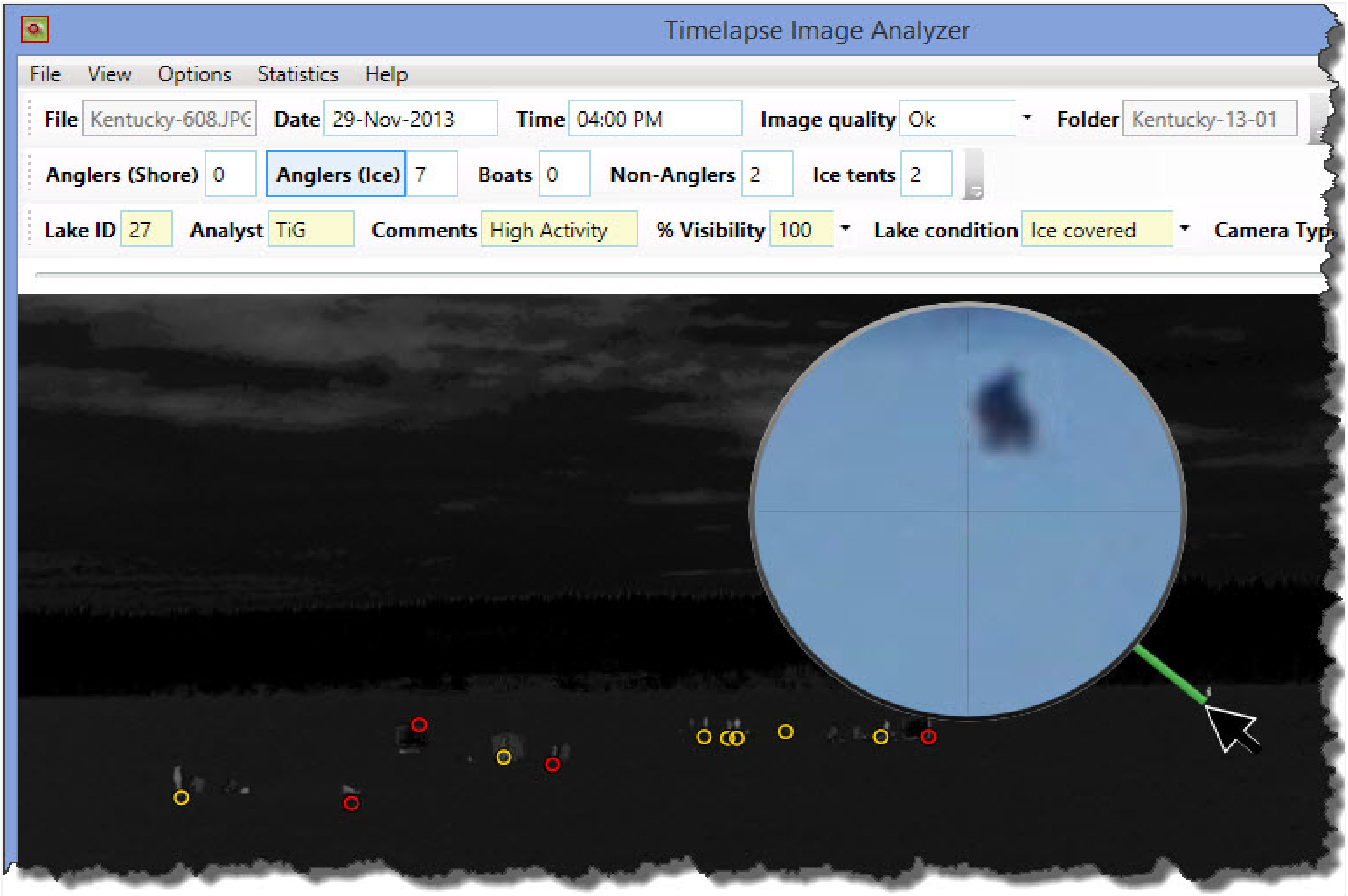

Figure 6. The Timelapse image differencing feature, where the analyst is inspecting the enhanced grey spot next to the cursor (see Figure 5 for how this spot appears in the original image).

TIMELAPSE eases search via several strategies. The first leverages magnification via a *magnifying glass* or *pan/zoom*, both which allow the analyst to search for and identify features by magnifying image areas. A magnifying glass (illustrated in Figure 5) magnifies the area immediately around the moving cursor (the magnification factor is adjustable). For example, the analyst may search for anglers by running the magnifying glass over the shoreline, or identify small features by magnifying them. Alternately, the analyst can zoom into any region of the image using the mouse scroll wheel, and pan around by dragging the image. While the magnifying glass is faster, pan/zoom magnifies more of the scene. In all cases, magnification is fluid and performed in real time.

The second strategy leverages *rapid image switching.* Because anglers move around, their position will change over time. Analysts exploited this in their search by switching rapidly between surrounding images, which makes even small changes ‘pop out’ as an animation. Yet many photo viewers reset the zoom level when switching between images, which meant that analyst could only use image switching when fully zoomed out. TIMELAPSE remedies this by keeping the magnifying glass location and the pan/zoom settings constant as the analyst switches images. Thus successive images are shown at exactly the same zoom level and centered on exactly the same location, where image features appear in exactly the same spot. Keyboard shortcuts (left/right arrow keys) make image switching rapid.

The third strategy implements *image differencing* as an alternative to image switching, where TIMELAPSE creates a composite image that highlights visual differences betSween the surrounding images to the current image (see Figure 6). Internally, TIMELAPSE compares pixels across images; if pixel values differ by a given threshold (to remove noise), that pixel is drawn in grey indicating the degree of difference, otherwise black. As Figure 6 illustrates, differences ‘pop out’ as bright spots. Yet because such spots can also be caused by non-interesting differences (e.g., moving shadows, objects moving in the wind), the analyst must still inspect those spots by examining the unaltered image in the magnifying glass (Figure 6).

Entering data.

Analyst originally entered data on a spreadsheet. This proved error-prone, especially when they mis-entered data into the wrong row or column. It was also tedious: because analysts preferred keeping the image viewer maximized in order to see image details, they had to repeatedly switch back and forth between the viewer and the spreadsheet, sometimes as many as 30+ times per image. They reported problems tracking what was counted on images with many entities in it. They also said data entry was tiresome when recording information that changed little across multiple images.

TIMELAPSE simplifies data entry. First, as seen in Figure 5, it obviates window switching, as its single interface combines both the image (bottom) and the data fields to be filled in (top). The analyst then enters data during image examination.

Second, TIMELAPSE minimizes typing. Only the data of the type *note* requires typing (e.g., ‘Lake ID’, ‘Analyst’, and ‘Comments’ in Figure 5). In contrast, *fixed choice* data types provides a drop down menu displaying a list of valid choices as specified by the biologist in the data template (e.g., the ‘Camera Type’ menu in Figure 5). The *count* data type implement a special interface that simplifies and adds accountability to the counting process. To count a particular entity, the specialist selects the data field (e.g., ‘Anglers (Shore)’, ‘Anglers (Ice)’, etc. in Figure 5), and then clicks on those parts of the image where that entity appears. A colored-coded marker is drawn atop that location and increments the count in that data field. The analyst can also remove that marker to decrement the count. If the analyst hovers over that mark, its associated data field is highlighted and a description displayed: this allows double-checking. Markers are especially important for minimizing errors on busy images, as the marks make it easy to see what was or was not counted.

Third, TIMELAPSE simplifies entering of data that changes little across multiple images (e.g., environmental changes over the course of a season, such the “Ice covered” Lake Condition in Figure 5). The ‘Copy Previous Values’ button (Figure 5**,** top right**)** applies to data fields marked *copyable* in the data template (Figure 4), where it copies data from the previous image to the currently viewed image. Each data field also has a context menu that allows one to copy or propagate values across images in various ways, e.g., across all images in the set, or to back-fill empty fields from various points. Using these techniques, the specialist only needs to enter data when differences occur, rather than image by image.

Finally**,** we noted that analyzers wanted to deal with particular types of images in bulk, such as nighttime shots. TIMELAPSE provides *image filters* that show only the subset of images matching a particular attribute. One filter, for example, displays only very dark images (specified as ‘Dark’ in the ‘Image Quality’ field; see Figure 5). Using that filter, the analyst can quickly scan them to verify that they are nighttime shots (which comprises over one-third of all currently collected images), and then propagate a single field across all those images to mark them as such e.g., %Visibility = 0. Another filter captures corrupted images, which can be dealt with in a similar manner. A third filter shows only non-dark and non-corrupted images, which will be the primary focus of the analysis. With these filters, the specialist can quickly deal with non-important images. We are currently modifying TIMELAPSE to create custom filters (e.g., ‘*images with angler counts > 0*’).

Correcting common errors.

Our analysis revealed several sources of errors that – while somewhat frequent – are painful to correct. One example is date and time errors caused by mistakes in initial camera set-up, or introduced due to changes between daylight saving and standard time, or from camera limitations. When this occurs, analyzers have to change the data and time fields manually across almost all images. To relieve this burden, TIMELAPSE allows the analyst to bulk-correct these errors (using dialog boxes) by: adding a correction value to all dates (which handles incorrectly initialized cameras); changing the time from a certain point onwards (which handles daylight saving/standard time issues); or by specifying a starting time that propagates by set intervals across all images (useful when dates and times cannot be automatically determined).

Validating data.

After the analysts complete their task, the data needs to be checked and validated. Beyond data accuracy, validation also allowed the biologist to gauge the abilities of the analyst. This was difficult to do in the original process, due to the separation of the spreadsheet data from the source images, and because it was unclear what the analyst had actually counted on each image.

TIMELAPSE eases validation, where biologists can efficiently spot check or thoroughly validate the data collected against each image. Because data is associated with each image, the biologist can quickly navigate through images to see what was counted, and to scan for anomalies. Corrections could be done on the spot. Because markers indicate what was counted, the biologist can check what entities were counted and how they were identified, and review the image to see if any entities were missed.

Discussion

TIMELAPSE has been in active use for almost four years, where Small Lakes Fisheries BC technicians and biologists have used it to classify over a million images from quite different lakes. Calculating angling effort estimates from image data is now routine practice. Compared to the original image analysis method, the improvements afforded by TIMELAPSE have been dramatic. First, the time to analyze images has been cut by ~60%, resulting in a marked cost savings. Second, the quality of data returned is far more reliable and standardized. Formatting errors are almost eliminated, and validations and corrections (if any) are fast to do. As a result, data can be uploaded directly into provincial databases and/or used almost immediately for estimating angling effort. High data quality also led to additional time and cost savings, e.g., data is rarely thrown out, and biologists do not have to spend valuable time correcting large flaws in data sets. Third, the system has been well received by technicians. While image analysis is still labor intensive, they feel that they are working efficiently.

We foresee how TIMELAPSE may further reduce costs by affording citizen science, i.e., the use of volunteers rather than fisheries technicians as analysts (Silverton, 2009). First, TIMELAPSE simplifies training. Volunteers can be given pre-populated image sets, where they would use TIMELAPSE to review how experts had identified image features. Volunteers can also be given training sets, where they would produce data that they can compared against expert results to see where they erred. Second, using crowdsourcing, image sets can be sent to volunteers who can analyze them at their leisure. Third, data can be further validated by sending the same image set to multiple citizen scientists. The system could automatically flag differences between these data sets, where it would present anomalies (rather than the entire data set) for review by the biologist.

While our own interests lie in small lakes, strategies such as those found in TIMELAPSE can be exploited in a broader range of projects employing remote cameras. Examples include: resource use across rivers, marinas, and passages; access ramp activity; parks use, facilities use; fishing demographics; boat traffic; and counting other human activities in the area such as biking or hiking. Indeed, its use goes beyond fishing: TIMELAPSE is currently being used by wildlife biologists to track wildlife and human use in national parks and other sensitive areas, where millions of images have already been analyzed.

Availability

TIMELAPSE software, installation instructions, tutorial documentation (including example image and data template files), and mailing list information are freely available at http://saul.cpsc.ucalgary.ca/timelapse. Documented source code is included, where software developers can modify or enhance its behavior if needed.

Acknowledgements

Funding was provided in part by NSERC and in part by BC's Small Lakes Management and conservation Initiative. We thank our analysts, who provided valuable insights into their analysis process.

References

Greenberg, Saul. (2013) TIMELAPSE Image Analyser: Software and documentation. Last modified November 20, 2013, http://saul.cpsc.ucalgary.ca/timelapse.

Greenberg, Saul and Theresa Godin. (2012) "TIMELAPSE Image Analysis Manual." Research Report 2012-1028-11, Dept. Computer Science, University of Calgary, Alberta, Canada, Last modified March, 2012, http://dspace.ucalgary.ca/handle/1880/49169

Holtzblatt, Karen., Jessamyn Wendell and Shelley Wood, *Rapid Contextual Design*. Morgan Kaufmann, 2004.

Kays, Roland and Keith Slauson "Remote Cameras", in *Noninvasive Survey Methods for Carnivores*, ed. Robert Long et. al., Island Press, 2008.

Newton, Eric., van Poorten, Brent., Godin, Theresa., Clarke, Adrian., Greenberg, Saul. and Post, John. (2013) "Using Cameras to Remotely Measure Angling Effort on Small Lakes," *66th Canadian Conference for Fisheries Research - CCFFR'13,*Windsor, Ontario, January 3-5, 2013.

O'Connell, Allan, James Nichols, K. Ullas Karanth (eds), *Camera Traps in Animal Ecology: Methods and Analysis*, Tokyo, Dordrecht, London, Heidelberg, New York: Springer. (2011).

Parkinson, E.A., J. Berkowitz, and C.J. Bull. ''Sample size requirements for detecting changes in some fisheries statistics from small trout lakes.'' *N. Am. J. of Fish. Mgmt.* 8:181-190 (1988).

van Poorten, Brent, H. Ward. and C. Walters, "Calculating Recreational Fishing Effort from Timelapse Cameras Using a Hierarchical Bayesian Model". b.vanpoorten@fisheries.ubc.ca University of British Columbia. In preparation.

Silvertown, J. "A New Dawn for Citizen Science", *Trends in Ecology & Evolution* 24:467–471 September, (2009).

Tredger, C. Dave, ‘’Estimation of Angler Effort Using Index Boat Counts.’’ *Fisheries Technical Circular No 94.* Ministry of Environment, Lands and Parks, Fisheries Branch, Conservation Section. 19pp.(1992)

Wiggins, A., G. Newman, R.D. Stevenson, and K. Crowston, “A Review of Mechanisms for Data Quality and Validation in Citizen Science”, *IEEE Seventh International Conference one-Science Workshops (eScienceW): Computing for Citizen Science*, (Stockhom). pp. 14-19, IEEE Press, DOI: 10.1109/eScienceW.2011.27, (2011)