

General comments:

The authors showcase the potential of near real time ablation monitoring in combination with long term mass balance data and modeling to assess the impacts of heat waves on Alpine glaciers. The manuscript explains an image analysis algorithm developed by the authors to automatically process images from their camera-based ablation measurement system. In a further step, the temporally highly resolved ablation data extracted from the images are used in a more general assessment of the extremely warm 2022 ablation season, which is then compared to a decadal regional average of glacier mass balance. This is an interesting and very timely contribution that shows the immense value of high resolution, real time glacier monitoring and data assimilation for impact assessments of extreme heat events. I have a few relatively minor comments/questions that I am sure can be addressed. I look forward to seeing this work published in TC.

Mainly, I would like some more detail on the image analysis algorithm, particularly in the discussion section. The methods and results go into a fair amount of detail on the algorithm, while the discussion focuses on the 2022 season in the regional context and related uncertainty analysis. Since it is a central part of this study, I would like to see some discussion of the automated image analysis. For example, I would be interested in more context on why template matching was chosen as opposed to other feature detection algorithms that might potentially handle rotation and scaling variations better. I would also be interested in learning more about whether/how changes in illumination affect the template matching.

The authors commendably produced a “read the docs” documentation for their template matching code. I found the examples of good and bad template images instructive. In my opinion it would be worth making these easier to find by putting them in the supplement or referring the reader more directly to the code documentation. The section on “known issues” (<https://rtgmc.readthedocs.io/en/latest/outlook.html>) answered some additional questions I had after reading the manuscript and some of this content might be added to the discussion. The fact that changes of more than 4 cm between images cannot be measured by the algorithm seems like a significant limitation in terms of applying the algorithm to similar use cases with slightly different measurement setups, e.g. with a lower temporal resolution and thus greater ablation between consecutive images. Could this be solved in the future by adapting the algorithm or applying some other, perhaps CNN based feature detection procedure? I understand that this is not supposed to be a computer vision paper but I think that giving a “glaciological perspective” on some common problems of feature detection and possible solutions would be a valuable addition to the discussion.

Specific comments:

Terminology:

- “stake displacement” → I stumbled over this because I associate “displacement” with stakes changing position due to ice movement. Perhaps an alternative term could be found (something like stake emergence?), otherwise I suggest clearly defining your usage of “displacement” somewhere at the beginning of the methods section.
- State somewhere what exactly you mean by “automated readings”, “visual readings”, “in situ readings”, “manual readings”. It is mostly clear from the text but having it in writing

would help the reader. “manual readings” is used interchangeably (I think?) with “visual readings”. Consistent wording would be better.

Abstract

“Compared to the average course of the past decade, the 25 days of heat waves in 2022 caused a glacier mass loss that corresponds to 56% of the overall mass loss experienced on average during summers 2010-2020, demonstrating the relevance of heat waves for seasonal melt.” → Consider rephrasing. I was initially unsure if “average course of the past decade” and “average during summers 2010-2020” refer to different averages. Perhaps just start the sentence with “the 25 days...” to shorten the sentence and avoid the repetition.

Introduction

L57 “In this study, we present an approach for the automated reading of the color-coded ablation stakes proposed by Landmann et al. (2021), which allows deriving daily point mass balances via the direct glaciological method” → consider rephrasing to more clearly distinguish between what Landmann did (developed camera/stake system) and what you did (automation). This sentence sounds like you “present an approach for something proposed by Landmann”.

Methods

L105 “Here, we assume that the line intersecting the highest number of matches corresponds to the stake” → when you later mention projecting onto the stake axis, do you mean you are projecting onto this intersecting line, or do you detect the stake axis in another, additional step? If so, how? If not, clarify that “stake axis” is the same as this line.

Fig3, 4 and related text: How do you deal with rotation of the image in the template? Is the rotation not large enough to matter for small time steps, or do you somehow correct for this? A reference to the video supplement somewhere in this section would also be helpful.

Fig 4: if possible: add arrows or similar to the 1 count occurrences in panel d to show which matches in panel c these represent. ie, somehow mark “erroneous” distances between matches in panel c and link them with panel d to make it easier for the reader to understand the concept.

L130 “Here, $d_{ref,px}$ is the **most frequent distance between** all possible tape combinations in an image, and $d_{ref,cm} = 4 \text{ cm}$ is the reference distance between two tapes. In the example of Figure 4 a, **the average distance between the tapes** corresponds to 42.7 px, resulting in a conversion factor of $c = 0.094 \text{ cm px}^{-1}$. . . The impact of such differences is reduced by taking the average pixel distance”

→ is it the most frequent distance or the average distance? Unclear.

L140 “During summer snowfall, the algorithm is not able to capture accumulation, and the stake displacement, equal to zero, is assumed as the ablation.”

→ Consider adding a note on why it does not work. Camera does not move and/or gets covered?

Figure 5: I find the inset arrangement of the panels counterintuitive. Perhaps this figure would work better if the panels are next to each other?

L209 “We attribute modelled daily mass balance of the closest series to every glacier of the most recent Swiss glacier inventory (Linsbauer et al., 2021) and match the daily time series to the glacier-specific annual mass balance (see above) by equally attributing the misfit to all days of the summer season (June-August).”

→ I struggle to understand this sentence. Does “closest” refer to spatial proximity of the 20 glaciers with GLAMOS data to all other glaciers? Does “see above” refer to the bias correction in the previous paragraphs? Is “misfit” the same as the bias explained earlier? If so, consider rephrasing to use the same terminology. If not, explain or otherwise clarify/rephrase.

Results

L250 and following paragraph → this explanation of how heat waves are defined might be moved to the methods section and could use some more detailed explanation. You cite a study by Hutter et al, who in turn cite Kysely (2002) for their definition of a heat wave. It would seem appropriate to cite Kysely here. Note that Kysely’s (and Hutter’s) usage is a little more complicated than “consecutive days above 30°C” and allows for hot periods separated by short, minor drops in temperature to be classified as one heat wave. It sounds like you did not use this part of their definition? They apply their heat wave definition to single locations and discuss heat waves in terms of their impact on excess mortality, choosing thresholds related to human well being rather than purely to climatological extremes. Your stations have a noticeable amount of altitudinal range and are situated on different sides of the Alps. Why did you choose these particular stations and a very broad, regional averaging process, rather than a more local assessment of temperatures? Given the abundance of weather stations in Switzerland, I assume meteorological data from locations much closer to your study sites would be available. Why not use stations at higher elevations that might better reflect the meteorological conditions at your study sites? You go on to state that extreme melt occurred outside of heat wave periods as per your definition - doesn’t this indicate that some other definition of “heat wave” (e.g. using climatological anomalies) might be more appropriate for this application? I have no major objections to your heat wave definition in principle but I would like to see more explanation on why you decided to do it this way, particularly since “heat wave” is in the title of the paper and the concept is central to many of the arguments you present.

Kysely J (2002) Temporal fluctuations in heat waves at Prague-Klementinum, the Czech Republic, from 1901– 1997, and their relationships to atmospheric circulation. Int J Climatol 22: 33–50

L266 The temperatures recorded during June and September 2022 were significantly higher than in the past decade and were thus able to cause extreme melt, but were not high enough to be also categorized as heat waves.

→ Rephrase to better distinguish when the June temps were a heat wave and when they were not, e.g. by giving dates for the heat wave. This is shown in fig 9 but the above sentence is hard to understand.

Fig 9 : what is the shading around the red and blue lines? uncertainty? Please add to the caption.

Discussion:

As mentioned above, I think a brief subsection discussing issues related to the template matching algorithm and how the algorithm might be adapted or improved would be appropriate. Also consider adding a few comments on how different definitions of heat waves might affect the overall calculations of storage change.

Typos

BE / AE spelling not always consistent (e.g. modeling / modelling)

L24 (Patro et al., 2018; Schaepli et al., 2019)) → remove extra parentheses

L26

“Despite glacier mass balance has been studied extensively with remote sensing (Bamber and Rivera, 2007), in-situ observations (Zemp et al., 2009), and modelling approaches (Hock, 2005; Hock et al., 2019), daily-scale mass balance variations remain mostly unexplored.” → “despite” should be followed by a noun or gerund, consider replacing “despite” with “although” or rephrasing.

L54 Because these novel methods **allows** measuring → allow

2 Study **site** and field data → study sites? (plural)

L101 The choice of a low correlation threshold **has also** some drawbacks. → also has (position adverb between subject and verb)

L151 By comparing the **algorithms** outcomes with the visual image readings, the daily errors of the automated approach can be computed. By comparing the **algorithms'** outcomes with the in-situ observations, the seasonal mean-absolute deviation (MAD) is derived. → algorithm's

L231 Of note is the fact that the daily deviations are mostly negative, which might suggest that the automated algorithm **overestimate** melt when compared to the visual readings. → overestimates