

S1 Snow profile distortions

When comparing snow profiles, avalanche forecasters automatically account for layer variations such as the same layers not being located at exactly the same depths or the same layers not having exactly the same characteristics (e.g., thickness, hardness, grain type, etc.). We refer to these differences as *distortions*. Numerical snow profile analysis tools that try to emulate the human snow profile assessment process need to be invariant to these distortions. In this section, we provide a brief overview of the type of distortions that snow profile analysis tools need to be able to handle and illustrate them with idealized, synthetic profiles. Interested readers are referred to Batista et al. (2013) and Paparrizos and Gravano (2015) for detailed reviews on invariances in time series analysis techniques.

Systematically different snow deposition (uniform scaling invariance; Fig. S1a): Two snow profiles may have different total snow heights due to consistent differences in the amounts of snow received at the two locations throughout the winter. In this case, the two snow profiles can be scaled (i.e., *Uniform Scaling*) to optimally align the layers.

Unsystematically different snow deposition (local scaling invariance; Fig. S1b): Even after appropriate uniform scaling of the profiles, layers may still be misaligned due to random differences in the amounts of snow that was received in individual storms. In this case, the two depth grids also need to be aligned *locally*.

Missing layers (occlusion invariance; Fig. S1c): It is not uncommon that layers that are present in one profile are missing in an otherwise similar profile due to specific local conditions (e.g., surface hoar layer that is only present in sheltered areas, or localized snow showers). Missing layers can either be located at the snow surface, within the snowpack, or at the interface between ground and snow. Numeric algorithms processing snow profiles need to be able to recognize missing layers and skip them when matching the adjacent layers.

Shifted layer characteristics (amplitude and offset invariance Fig. S1d): Differences in local conditions can also result in differences in the profiles of layer characteristics. For example, the hand hardness profiles of two snow profiles might exhibit a similar overall pattern but be shifted and/or amplified to (systematically) higher values in one profile.

Ideally, numerical algorithms developed for snow profile analysis are able to cope with all of the distortions at the same time (Fig. S1e).

S2 Derivation of hyperparameter and alignment approach recommendations

Aligning snow profiles with the DTW algorithm presented in this paper requires the specification of the hyperparameters *averaging weights* w_g and w_h , as well as the *window size* ε . In addition, users can select between different alignment approaches (e.g., combined bottom-up/top-down). In this section, we derive the recommended hyperparameter values and alignment approach presented in the main body of this article using two simulation experiments.

S2.1 Experiment 1: Grid hyperparameter values

The first experiment is a grid search to determine the optimal values for the hyperparameters *averaging weights* w_g , w_h and *window size* ε . For this experiment, we chose 3020 snow profile pairs from our operational 2018/19 winter dataset of modeled snow profiles in British Columbia. To ensure that the analysis dataset included both thin early season as well as mature mid-season snowpacks from different snow climates, we randomly drew 20 pairs from our operational dataset for each day between November 1 2018 and April 1 2019 across a range of regions. To include profile pairs with different levels of similarities in the analysis dataset, half of the profile pairs were selected to be less than 50 km apart, whereas the other half of the profile pairs were selected to be between 50 – 130 km apart. To further ensure meaningful pairs, all profiles were taken from the treeline elevation band, such that the intra-pair elevation difference was less than 135 m for the 75% of the pairs.

To determine the optimal hyperparameter values, we aligned each of the snow profile pairs with varying hyperparameter choices and determined the optimal choice based on the independent snow profile similarity measure Φ (see Sect. 4 for definition). We varied w_g from 0 – 1 in steps of 0.1 (w_h varied accordingly following $w_h = 1 - w_g$) and ranged ε within

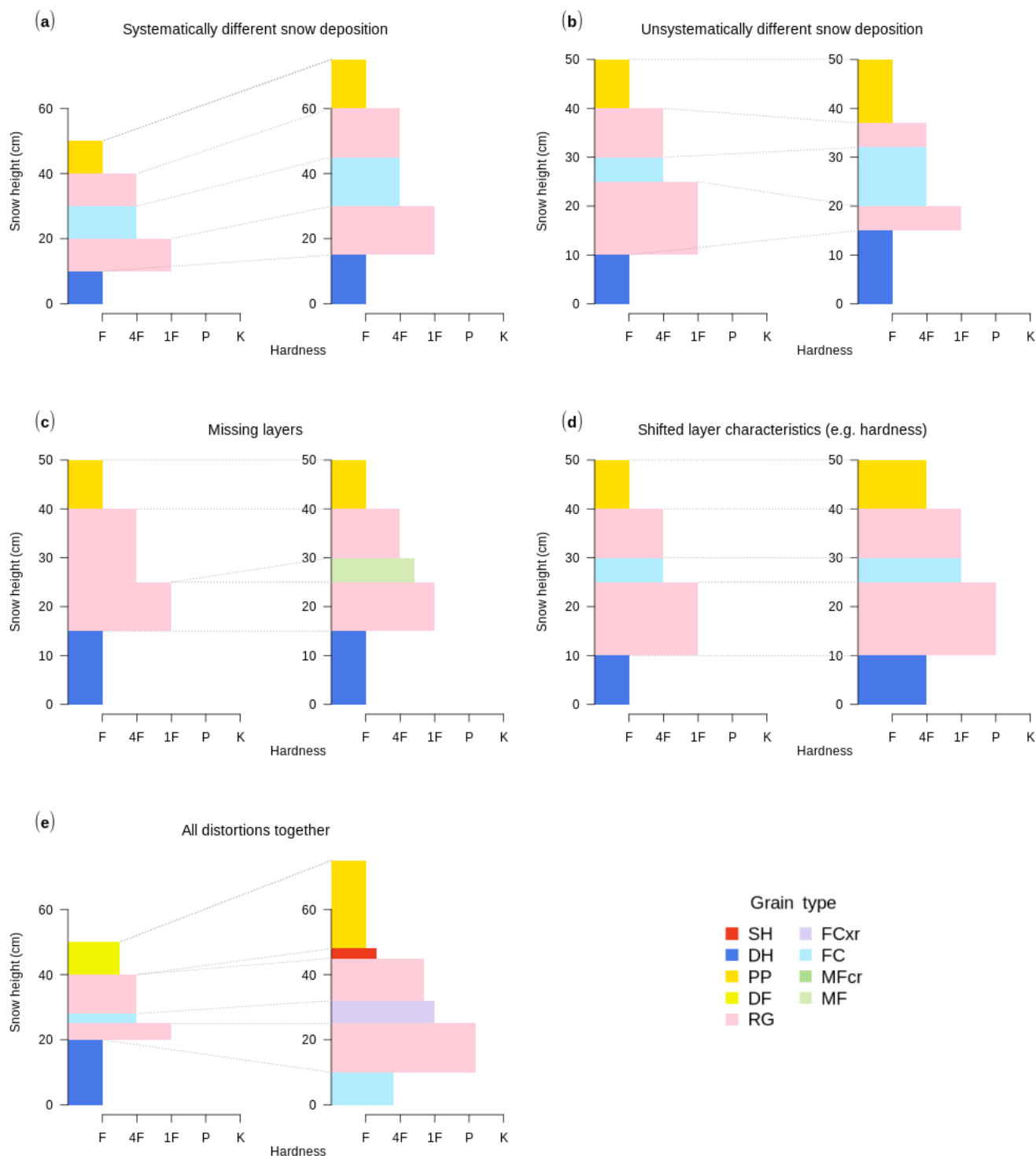


Figure S1. Idealized snow profile distortions. Ideally, numerical algorithms developed for snow profile analysis are able to cope with all of the distortions both individually and at the same time.

$\{0.1, 0.15, 0.25, 0.5\}$. Each profile pair was aligned bottom-up and top-down with all of the different hyperparameter settings, and the similarity measure Φ was calculated for every alignment. In total, this procedure resulted in 133k computed alignments.

Given that our sample consists of snow profile pairs that are fairly similar, we can assume that higher values of Φ represent higher quality alignments. Hence, the optimal choice of parameter settings for a snow profile pair should yield the highest similarity measure Φ_p^{max} . Note, that Φ_p^{max} for a profile pair can be obtained not only with one but multiple different hyperparameter settings. To make the quality of the alignments comparable across all profile pairs, we normalized the similarity measure for each pair with $\tilde{\Phi}_p = \Phi_p / \Phi_p^{max}$.

The results of our simulation experiment show that when looking at the distributions of *all* computed alignments (Fig. S2a), the sensitivity of the alignments with respect to w_g is only very limited in the $[0.5, 1.0]$ range. However, an examination of the proportions of *optimal* alignments (i.e. $\tilde{\Phi} = 1$) shows that they increase steadily as w_g is increased from 0 to 0.6 where the growth flattens out (Fig. S2b). While our results show that the highest proportion of *optimal* alignments is achieved by setting w_g to 1 (i.e., not including hardness information in the matching at all), this is not a meaningful choice as hardness information is critical for solving ties for layers that consist of identical grain types. This over-prioritization of grain type when determining the optimal choice for w_g is likely an artefact of the definition of our the similarity measure Φ , which focuses more on grain type than hardness. Hence, we suggest to set $w_g \in [0.6, 0.9]$ for optimal alignments.

At first glance, alignment performance seems to be best for rather large window sizes ε (Fig. S2c), which contradicts general expert recommendations from DTW research (e.g., Ratanamahatana and Keogh, 2004). We therefore conducted a second series of alignments with w_g fixed at 0.8 and a more complete range of ε values ($\{0.01, 0.05, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 1.00\}$). Using the highest similarity measure Φ_p^{max} of each snow profile pair as a reference, we then determined whether the window size was too small, optimal, or too large for each alignment. A window size was considered to be too small, for example, when Φ was smaller than Φ_p^{max} , and Φ_p^{max} occurred at a higher ε value. Fig. S2d nicely illustrates how the window size is too small for the vast majority of alignments at $\varepsilon = 0.01$ and how the proportion of optimal window sizes steadily increases up to $\varepsilon = 0.3$. After that, the proportion of optimal window sizes remains roughly constant as the windows that are too small are replaced by windows that are too large. The results of the experiment also revealed that the optimal window size ε varied across profile pairs (not shown), but typically stayed within $[0.15, 0.35]$. While this issue could be addressed by aligning all profile pairs with multiple window sizes and picking the optimal one on a case-by-case basis, this solution comes at a high computational cost. Instead, we recommend a window size of $\varepsilon = 0.3$ as the general default value for snow profile alignments.

S2.2 Experiment 2: Alignment approaches

To identify the optimal alignment approach, we conducted a second simulation experiment, for which we drew 1520 snow profile pairs (10 per day) from our operational 2018/19 snow profile dataset in the same fashion as for Experiment 1. In this evaluation, we explored four different alignment approaches:

Lock-step measure (LSM) This approach rigidly compares the i th layer of one profile to the i th layer of another one without any alignment. Even though this approach is extremely simple and ignores the natural variability of the snowpack, it has been found to be surprisingly competitive against more sophisticated approaches in other fields (Keogh and Kasetty, 2002; Wang et al., 2013).

Scaled lock-step measure (scLSM) In this variation of the LSM approach, the profiles are rescaled to identical snow heights (see Sec. 3.2 for details), but the layers are still not aligned.

Constrained bottom-up DTW with multiple window sizes (cDTW-x ε) In this approach, the profiles are aligned with DTW *bottom-up* where the first snowpack layers near the ground need to be matched, but the matching of the surface layers is relaxed (i.e., open-end). This alignment is performed with multiple window sizes $\varepsilon \in \{0.01, 0.05, 0.1, 0.15, 0.3\}$, and the alignment with the highest similarity measure Φ is identified as the best alignment.

Constrained bottom-up and top-down DTW alignment with fixed window size (cDTW-BU/TD) In this case, we aligned each profile with DTW twice; one with a bottom-up approach where the alignment requirement for the surface layers is relaxed, and one with a top-down approach where the alignment requirement for the layers at the bottom of the snowpack

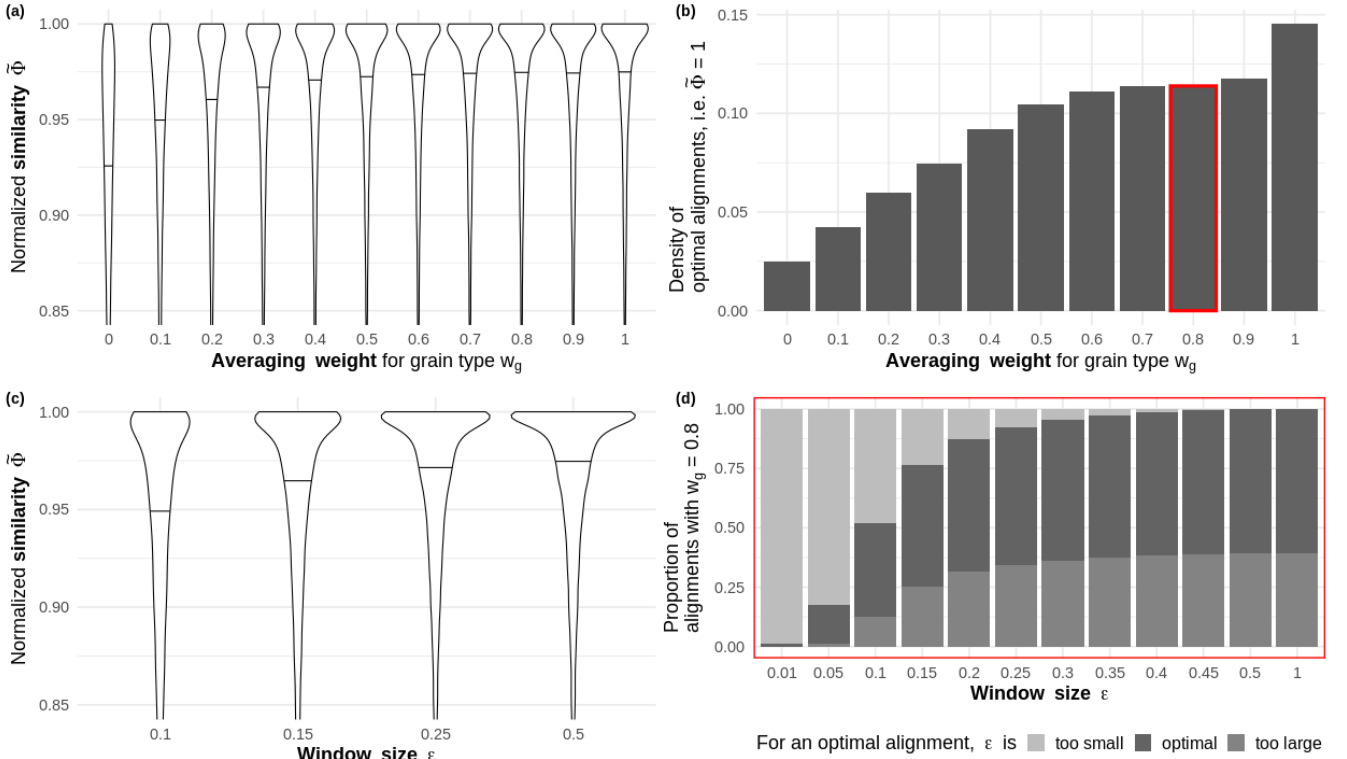


Figure S2. (a, c) Violin plots (including medians) of the normalized similarity measure $\tilde{\Phi}$ for varying hyperparameters averaging weight w_g and window size ε ; (b) the corresponding density of w_g for the optimal alignments; and (d) proportions of window sizes that are too small, optimal, or too large to yield optimal alignments, given $w_g = 0.8$.

are relaxed. For both alignments, we set the window size $\varepsilon = 0.3$ (cDTW-BU/TD). Similar to the third approach, the alignment with the higher similarity measure Φ is picked as the better alignment.

For each snow profile pair, we calculated the similarity measure Φ for every approach. To compare the similarity measures for the different alignment approaches across the entire dataset, we normalized the similarity measures for each snow profile pair with the similarity measure of the scLSM approach. Normalized similarity measures larger than 1 indicate that the alignment was better than with the scLSM approach, whereas values smaller than 1 represent worse alignments.

Our comparison revealed that the LSM approach on average yielded poorer results than the scLSM method (Fig. S3), which suggests that rescaling profiles to equal snow heights with a prescribed scaling factor improves the alignment of the layers. The comparison also highlighted that both DTW approaches clearly outperformed the lock-step approaches. cDTW- $x\varepsilon$ resulted in better alignments than scLSM in *every* case, cDTW-BU/TD did so in 99.9 % of the cases. However, we recommend the use of the cDTW-BU/TD approach over cDTW- $x\varepsilon$ due to its lower computational cost (2 versus 5 alignments for each profile pair).

In summary, we recommend the following default settings for the use of our algorithm for aligning snow profiles:

1. *Averaging weights* $w_g = 0.8$ [0.6, 0.9]
2. *Window size* $\varepsilon = 0.3$
3. *Alignment approach*: cDTW-BU/TD approach

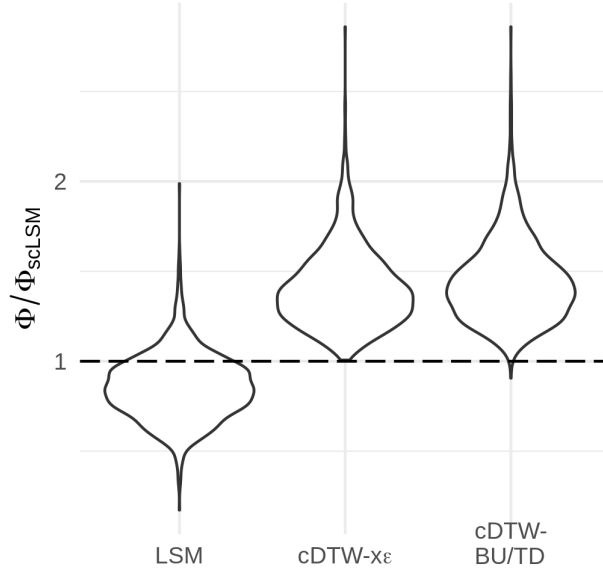


Figure S3. Violin plot of the normalized similarity scores for snow profile alignments carried out with different alignment approaches relative to the scaled lock-step approach (scLSM): (i) raw lock-step approach (LSM), (ii) constrained DTW approach with various window sizes (cDTW- $x\epsilon$), and (iii) a bottom-up and top-down constrained DTW approach with single window size (cDTW-BU/TD).

S3 Snow profile alignment examples

In this section, we present additional alignment examples to further illustrate the use range and behavior of our algorithm.

S3.1 Alignment of modeled versus human profiles

Simulated snow profiles typically contain much more details than human profiles. This is mainly because models can resolve small changes in the layer properties better than observers in the field, but also because human avalanche practitioners often *choose* to simplify layers they are not interested in or concerned about. Our algorithm is able to properly align simulated snow profiles with human profiles with different levels of detail.

Fig. S4 illustrates the alignment of a simulated profile with a highly idealized whiteboard profile, which summarizes only the most important features of the snowpack of an entire forecast region. Fig. S5 presents the alignment of a simulated profile with a detailed snow profile observation recorded by a human observer. The two examples clearly demonstrate that our alignment algorithm can handle the different amounts of detail between the modeled and the human profiles. The ability to computationally compare simulated and observed snow profiles creates new opportunities for evaluating the performance of numerical snowpack models across space in real time. Note, that both examples were aligned solely based on grain type and hardness information. For Fig. S4 the averaging weights were set to $w_g = 0.6$ and $w_h = 0.4$ because idealized whiteboard profiles have fewer grain type details than modeled profiles.

S3.2 Layer tracking

It is common practice among avalanche forecasters to track the evolution of snowpack layers of interest through time. That means that two profiles from the same location but different times need to be aligned. Sect. 3.5 of the main body provides recommendations for the use of our alignment algorithm for this type of application.

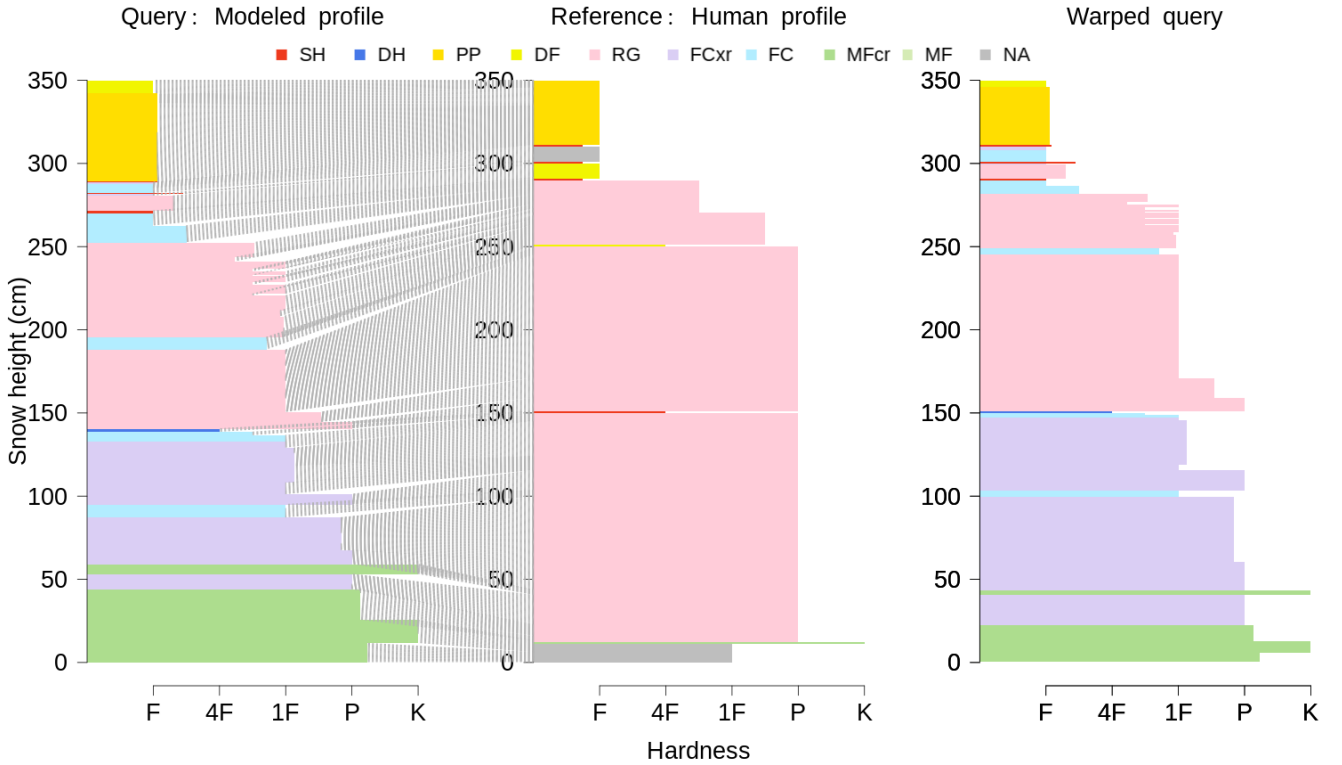


Figure S4. Alignment of a detailed simulated profile to an idealized, human whiteboard profile, which summarizes only the most important layers of an entire region.

To illustrate this type of use of our algorithm, we are aligning two simulated profiles from the same location but with a time lag of nine days (Figure S6; Profile 1: 2018-11-21; Profile 2: 2018-11-30). For reference, a thin surface hoar layer (SH) is labeled with its burial date (2018-11-13) in both profiles. During the week between the two profiles, the snowpack settled considerably, and the layer of new snow (DF) that overlay the labeled surface hoar in the earlier profile metamorphosed into a sequence of round grains (RG) and facets (FC). Additionally, a new storm added 50 cm of new snow on top of the profile. The alignment determined by the algorithm perfectly matches with this snowpack history.

For obvious reasons, an open-end, bottom-up alignment approach is best suited for layer tracking applications. Note that the alignment for this example was carried out with rescaled profiles and a window size of $\varepsilon = 1$. In this particular example, the perfect alignment is achieved with grain type and hardness information alone, and providing date information for key layers does not make any difference. However, in general the alignment performance for tracking layers will be better and more robust when at least few key layers are labelled with their burial date.

S3.3 Alignment of profiles that requires additional date information

Under some circumstances, it is impossible to conclusively align snowpack layers based on grain type and hardness information alone. This can be a challenge even for human experts. However, the issue can commonly be resolved by labeling one or several key layers with their burial date. We illustrate this challenge and the solution for it with the two snow profiles shown in Fig. S7.

Using only grain type and hardness information, the algorithm aligns the two profiles in a way that the dominant FCxr layers between 75 and 130 cm in the query profile Q are matched with the FCxr layers between 55 and 80 cm in the reference

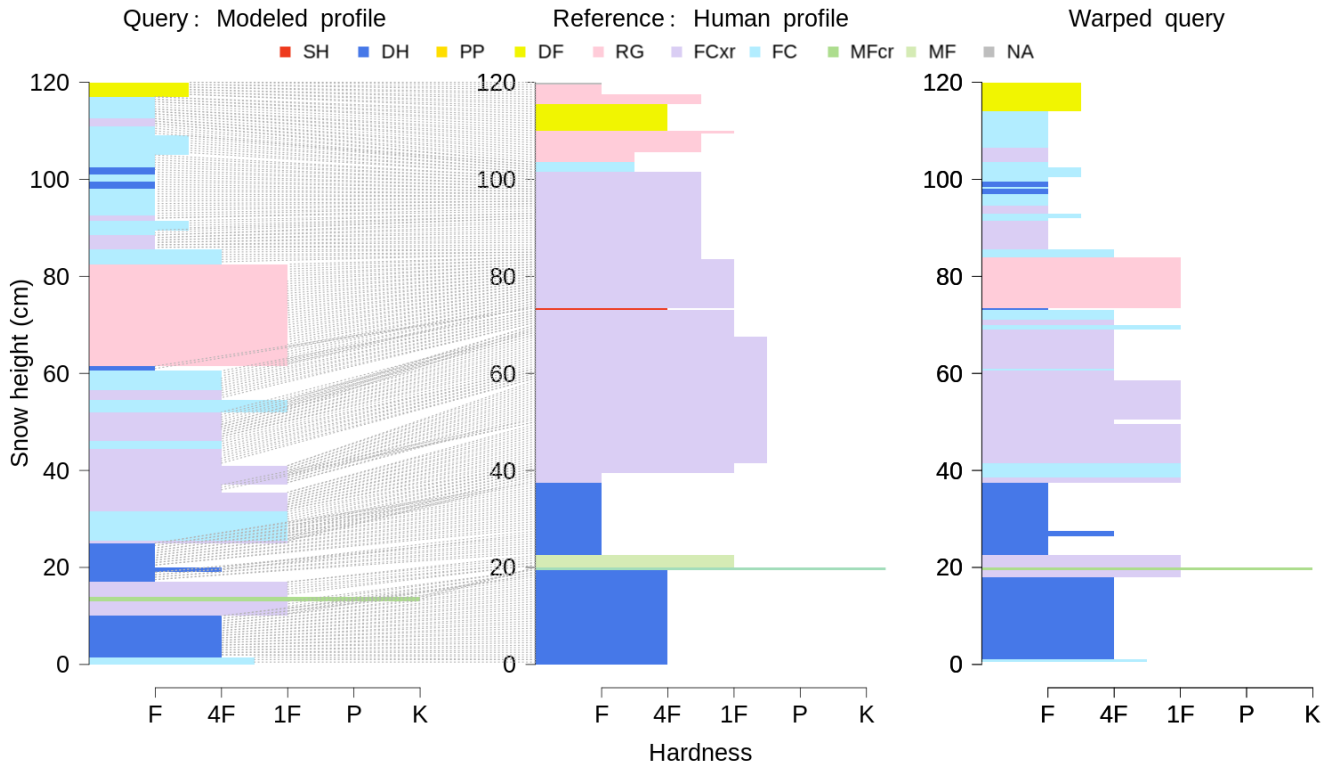


Figure S5. Alignment of a modeled profile to a detailed human snow profile observation.

profile R. However, human forecasters easily recognize that this is likely not a meaningful alignment due to the similarity between the three DH layers between 40 and 65 cm in the query profile Q and the FC layers between 55 and 85 cm in the reference profile R. The alignment algorithm fails to align the layers correctly for the following reasons:

1. The three weak layers have different grain types and hardnesses (i.e., DH–4F vs. FC–1F), and
2. The layer sequences *around* the three weak layers differ (i.e., FCxr above the weak layers, and FC between them vs. RG above the weak layers and FCxr between them).

Most of the time, these types of challenges are resolved by the preferential layer matching mechanism included in our alignment algorithm (see Sect. 3.3 for details). In the present case, however, the preferential layer matching is not sufficient to correctly align the profiles. Even though the correct weak layer matching is properly identified in the local cost matrix (yellow cells represent a low distance, or cost, and are thus matched more easily, Fig. S7a), the path of least resistance through the cost matrix avoids these matches. Labeling one of the three distinct weak layers with its burial date resolves the issue and yields the expected profile alignment (Fig. S7b). Adding the burial date information changes the structure of the local cost matrix in a way that anchors the warping path at the weak layer and results in more realistic matches in its vicinity.

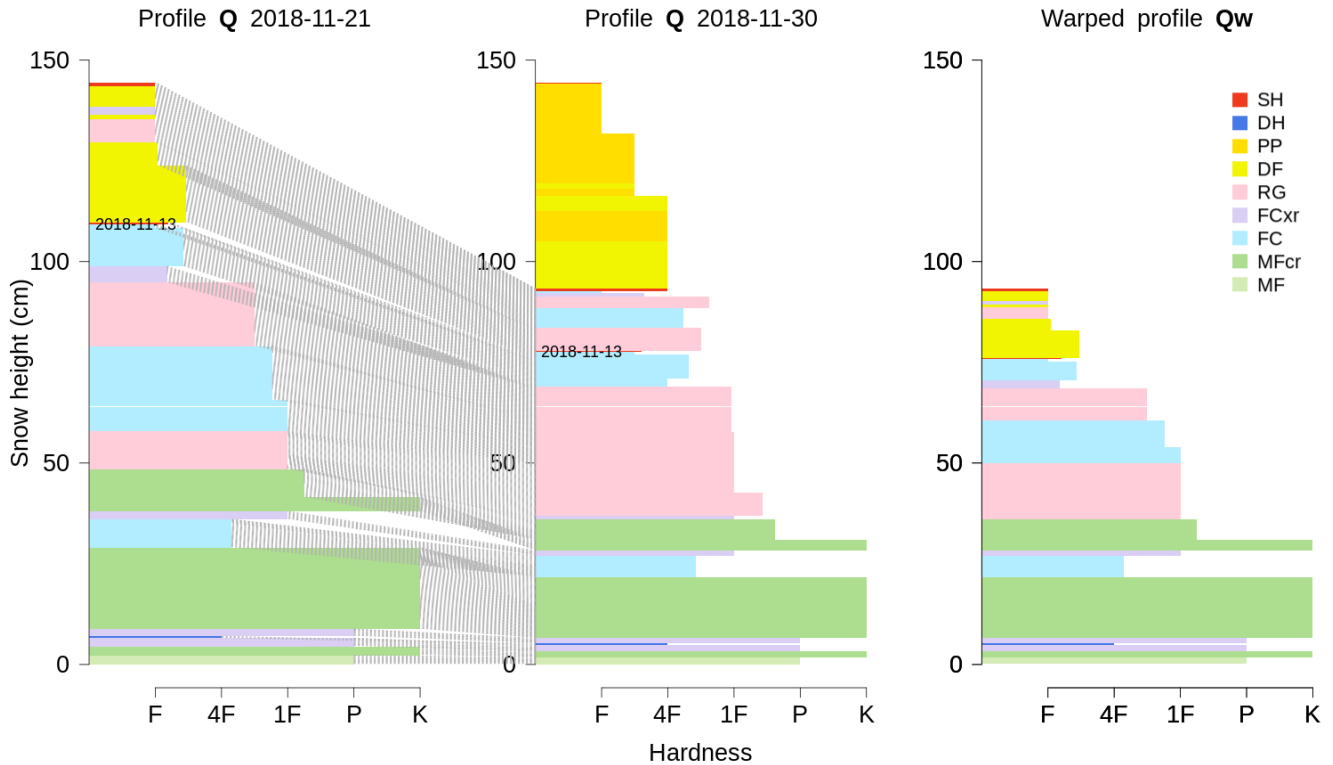


Figure S6. Alignment example of a layer tracking application: aligning profiles from the same location but a time lag of nine days to follow the evolution of layers through time.

References

- Batista, G. E., Keogh, E. J., Tataw, O. M., and De Souza, V. M.: CID: an efficient complexity-invariant distance for time series, *Data Min and Knowl Disc*, 28, 634–669, <https://doi.org/10.1007/s10618-013-0312-3>, 2013.
- Keogh, E. and Kasetty, S.: On the need for time series data mining benchmarks: A survey and empirical demonstration, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 7, pp. 102–111, Kluwer Academic Publishers, <https://doi.org/10.1023/A:1024988512476>, 2002.
- Paparrizos, J. and Gravano, L.: k-shape: Efficient and accurate clustering of time series, in: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 1855–1870, ACM, 2015.
- Ratanamahatana, C. A. and Keogh, E. J.: Everything you know about dynamic time warping is wrong, in: *Third workshop on mining temporal and sequential data*, vol. 32, Citeseer, 2004.
- Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., and Keogh, E. J.: Experimental comparison of representation methods and distance measures for time series data, *Data Min Knowl Disc*, 26, 275–309, <https://doi.org/10.1007/s10618-012-0250-5>, 2013.

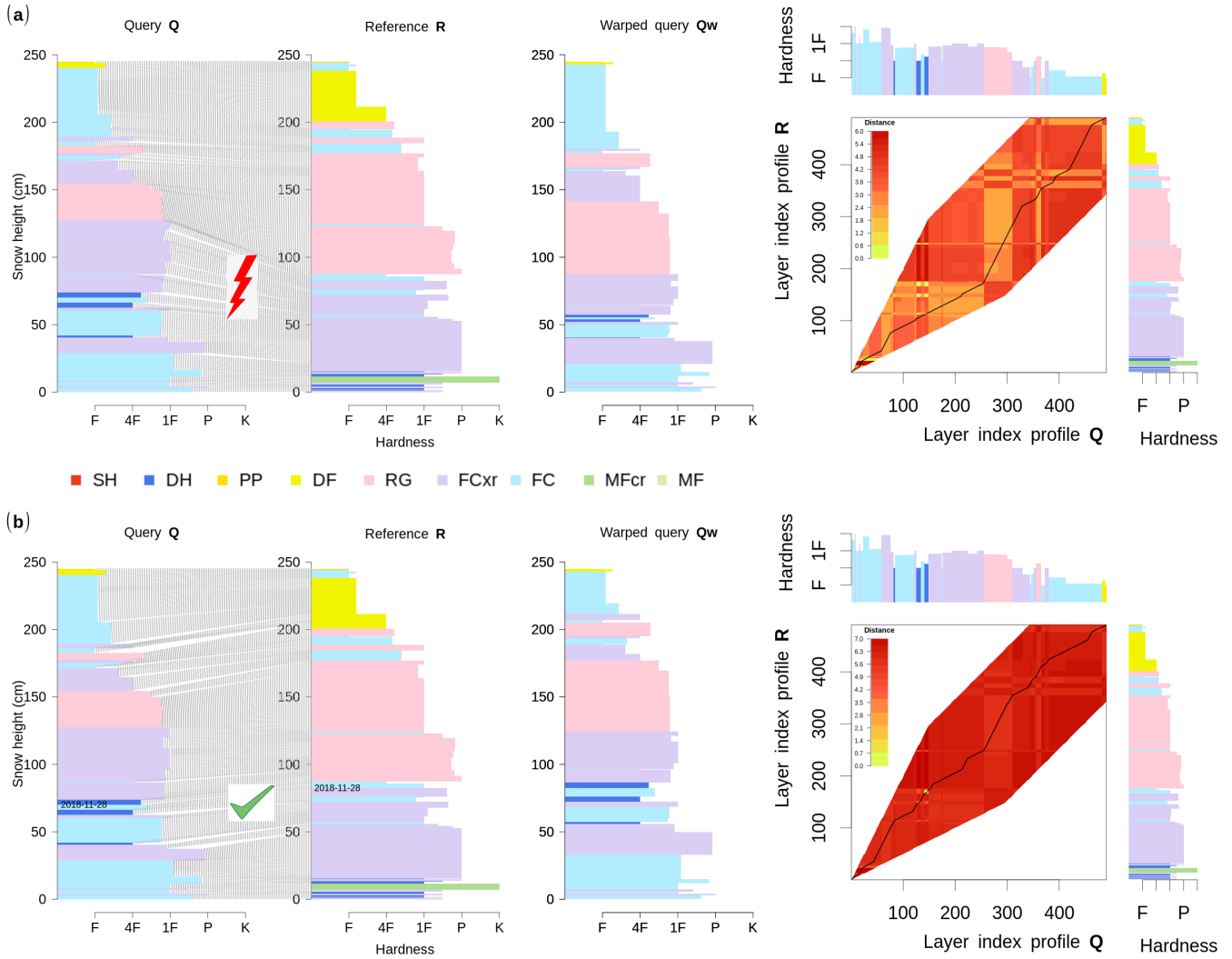


Figure S7. Alignment of the same two profiles Q and R (a) solely based on grain type and hardness, or (b) with one additional layer date information (i.e., 2018-11-28). In addition to the original and aligned profiles, the figure also shows the corresponding local cost matrices.