Dear Editor,
Dear Referees,

We would like to thank you for your very detailed comments on our discussion paper. We believe that your suggestions really helped us to improve the quality of the manuscript. Please find below our response to the Referee's comments (as published in the open discussion) and the revised version of the manuscript in track changes mode. Besides new model runs (sensitivity of temperature and precipitation, extended skill measures, additional table), some more literature has been added regarding hydrological forecasting, snow, tourism, and climatological aspects of predictability. In this way, we tried to best possibly address all suggestions and comments raised in the open discussion.

Kind regards,

Kristian Förster
(on behalf of all co-authors)

Table of contents

REPLY TO ANONYMOUS REVIEWER #1..2
REPLY TO ANONYMOUS REVIEWER #2..11
REVISED VERSION OF THE MANUSCRIPT..15

Reply to anonymous reviewer #1

Reviewer's comments are in italics

Summary
In this paper the authors analyze seasonal hydrological hindcasts following a dynamical approach. The hydrological model AWARE is forced with the output from two different GCMs (CFSv2, GloSea5) for the Alpine catchment Inn up to the gauge Kirchbichl. As main predictand the authors chose the snow accumulation, represented by the snow water equivalent, at the end of February with a lead-time of 4 months (forecast initialization in October the preceding year). Additionally this paper assesses the predictive skill of both GCM-based forecast with regard to the anomalies of basin-scale mean temperature and accumulated precipitation depth.
In my opinion, the manuscript fits well into this special issue and its content is relevant for publication in HESS. In particular I like the use of two different GCMs in combination with a water balance model and the analyses of a different hydrological predictand than flow. I recommend this paper to be published after the authors have addressed the following general and specific comments in order to further improve the manuscript.

We would like to thank anonymous referee #1 for his/her positive and constructive review of our discussion paper. Your comments and suggestions will help us in the process of revising our manuscript. Please find our detailed response below.

General comments
Overall, the paper is well written and well organized presenting interesting results. Nevertheless the authors should elaborate the following aspects: To facilitate readability of the manuscript you should explicitly indicate throughout the paper, if you're talking about meteorological seasonal forecast (e.g. used as input for a water balance model) or hydrological seasonal forecasts (output from a water balance model), as both cases are relevant in different parts of the text. I recommend to use the notation of the 4 model experiments (introduced in section 2.4) more consistently throughout the whole paper (including figures), for example clim. forecast should be CF-AWARE.

Yes, this is a good point. We carefully updated the manuscript with respect to the notation of model experiments. We are expecting to increase the readability in this way. Thank you for this recommendation.

I suggest re-arranging the validation of the hydrological model: in chapter 2.3 you solely assess the performance of the water balance model with regard to runoff, although the prediction of SWE is the focal point of this study. Therefore I suggest moving the SWE-related evaluation from section 3 (page 7) to section 2.3 and to focus primarily on SWE simulation.

We discussed moving the SWE evaluation to the model description section. At first sight, it would be a good alternative which seems to be straight-forward. However, this would increase the number of cross references in the manuscript, which, in our opinion, might not be helpful in the process of revising our manuscript. The model is calibrated and validated using runoff. This makes sense since the scales of the model and the observations match.

We didn't calibrate against SWE data. Moreover, the evaluation of SWE also includes a study of the representativeness of the climatological forecast (CF-AWARE). This type of forecasts is introduced in the section describing the model experiment (2.4) which follows the model description section (2.4). Moreover, it also depends on initial conditions. Therefore, we see the SWE comparison as part of the results section.
However, we agree that we should highlight the relevance for presenting these results here. We will add a piece of introducing text prior to this analysis:

"While the applicability of AWARE to reconstruct the water balance in terms of observed runoff time series was demonstrated in Sect 2.3, it is necessary to evaluate the model experiments HISTALP-AWARE and CF-AWARE with respect to SWE prior to the analyses of CM-based SWE forecasts."

Although you focus on climate model-based seasonal forecast you should add some more background information on the different approaches to create seasonal hydrological forecast (e.g. statistical methods vs. dynamical approaches) in the introduction. Furthermore I miss a better support of literature in the discussion (section 3), e.g. with regard to related studies of the Alpine region, e.g. Fundel, F., Jörg-Hess, S., and Zappa, M.: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. Hydrol. Earth Syst. Sci., 395-407, doi:10.5194/hess-17-395-2013, 2013 and with regard to the comparison of your skill measures (e.g. correlation) with other studies, e.g. Kim, H.M. , Webster, P. J., Curry, J. A.: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. Clim Dyn (2012) 39:2957–2973. doi 10.1007/s00382-012-1364-6 or Weisheimer, A., and Palmer, T.N.: On the reliability of seasonal climate forecasts. Journal of The Royal Society Interface 11, Heft 96, S. 20131162–20131162, 2014.

The revised version of the manuscript will include some more information about statistical and dynamical approaches in the introduction (please refer to our response to your specific comment on this topic below). In this context, we will also refer to the literature you have listed in your comment when discussing the results later in Section 3:

"Compared to findings reported in the literature, the results achieved in this study are promising given that the skill in Europe is generally found to be low. For instance, according to Weisheimer and Palmer (2014) the skill of DJF temperature is "marginally useful" using ECMWF's System4. The rating for DJF precipitation is even found to be "not useful" (cf., Fig. 5 in Weisheimer and Palmer, 2014). Similarly, Kim et al. (2012) found some skill in terms of correlation for wintertime temperature predictions using System4. However, their study also suggests low absolute correlation coefficients for precipitation forecasts and for both temperature and precipitation forecasts achieved using CFSv2. A direct comparison to the results presented in this study is not possible since GloSea5 was not addressed in these studies. Moreover, given that only one single catchment is considered, a ranking of models is beyond the scope of this article. The predictability for SWE detected in this study can be related to both some skill in precipitation prediction and previous findings found for the persistence in SWE predictions for smaller forecast horizons. For instance, in case of the alpine snow cover, Jörg-Hess et al. (2015) underline the persistence in SWE predictions at least up to a lag of two weeks."

You focused on "hydrological storages instead of instantaneous hydrological fluxes", which you call "a new aspect". For me it is not obvious . . . . . . if you already tried predicting fluxes for the melting season in the Alps? Is SWE indeed better predictable than the resulting flow? and . . . if SWE is a useful information / predictor for the stakeholder you mention (reservoir managers)? Is it as useful as flow forecasts or is it more like a fallback option?

We agree that snowmelt predictions are very useful for numerous stakeholders. Accordingly we added additional literature to the introduction. Yet, we didn't try to predict fluxes using CM-based meteorological forcing for the melting season because we expect the predictions to be less accurate than for SWE in winter. We added the following lines:

"We focus on the winter season as extratropical seasonal forecasts appear to have the highest skill in this season (e.g. Riddle et al 2013, Scaife et al 2014, Kang et al 2014). There are a number of reasons for this, including winter being the season when the stratosphere is active, which is known to affect predictions (e.g. Domeisen et al 2015, Scaife et al 2016, Butler et al 2016). The winter season also shows much stronger dynamical connections to the tropics, allowing high predictability of tropical rainfall (Kumar 2013) to be transmitted into the extratropics (Greatbatch et al 2012, Molteni et al 2015, Scaife et al 2017)."

In a different model experiment, spring forecasts have been studied based on initial conditions derived from CM-based seasonal model runs and HISTALP meteorological forcing (Förster et al., 2017). However, CM-based seasonal forecasts have only been addressed to study snow accumulation in the winter and simulations in spring are performed in the framework of a reverse-ESP experiment. Thus, fully dynamical seasonal forecasts of spring runoff have not been performed yet. However, in the discussion paper, we already addressed snowmelt forecasts in spring as an outlook for future research in the conclusions section (page 10, line 3-4). This topic could be addressed in the framework of a subsequent analysis.

For the hydropower stakeholders peak SWE, which is the starting point of the melt season, would be beneficial for reservoir inflow and flood forecasting. In the Alps peak SWE normally occurs in April or May. In a first step, we tried to make use of wintertime forecasts and the predicted SWE in February is viewed as a first guess for the catchment's state prior to snowmelt. Moreover, SWE is important variable for tourism in the Alps. This argumentation was also added to the introduction.

Please refer to our reply to your specific comment on the question whether predicting storages is actually more robust instead of predicting fluxes.

As you evaluate forecast skill for SWE together with areal precipitation and temperature, I think it's necessary to address the interaction of these parameters more explicitly: what is the relative contribution of precipitation compared to temperature on SWE in the Alpine basin at the end of February? I guess you have insights into this interaction, so please share it with the reader of your paper.

Evaluating the relative role of both temperature and precipitation in SWE forecasting is a question which has also been raised by reviewer #2. We will prepare a small model experiment in which climate model forcing is replaced by either climatological series of temperature or precipitation, respectively. This experiment gains insight into the question whether temperature or precipitation mainly contributes to the skill in SWE forecasts. The revised version of the manuscript will include a table of performance measures in order to highlight the contribution of temperature and precipitation, respectively. Please also refer to our response to the comments of Anonymous Reviewer #2.

Specific comments
I attach the specific comments as supplement.

Page 1, line 13: I agree that seasonal forecasting is a topic which became a focal point in hydrological forecasting in recent years. But the term "new" isn't adequate in my opinion, as the earliest references e.g. for ESP-based long-term forecasts have been published in the 1970s / 1980s (e.g. Day, G. (1985). "Extended Streamflow Forecasting Using NWSRFS." J. Wat. Res. Plan. Mgmt., 10.1061/(ASCE)0733-9496(1985)111:2(157), 157-170 or Twedt, T. M., J. C. Schaake, Jr., and E. L. Peck (1977). National Weather Service extended streamflow prediction. Proc., Western Snow Conference, 52 – 57.). Furthermore on page 2, line 8 you state that "seasonal outlooks [...] have been prepared for decades." – This is a contradiction to the term "new", too.

We agree that this argumentation in its present form is not consistent. Indeed, seasonal hydrological forecasting is a topic that is not so new as mentioned. Some additional literature was added to the introduction. However, using climate model based hydrological forecasting is new. Thus, we will revise the text regarding the historical perspective of its relevance. In the revised version, we will restrict the "new" aspect to climate model based seasonal predictions:

"Climate model (CM)-based seasonal predictions are an emerging new field in hydrology (e.g., Yuan et al., 2015; Svensson et al.,2015; Mackay et al., 2015)"

We will also add the references Day (1985) and Twedt et al. (1977) along with some other references, see below. Thank you for this important comment.

Page 2, line 12/13: This sentence is a bit confusing / misleading, because meteorological data is used for the ESP-approach, too. Furthermore I recommend extending your description of ESP / revESP in order to explain more clearly which components contribute to forecast skill in each approach.

You are right to say that this statement is misleading. We extended our explanations regarding ESP/reverse-ESP and added the references as suggested:

[…]
"This methodology is well known and referred to as Ensemble Streamflow Prediction (ESP, Wood and Lettenmaier, 2008). The development of this method goes back to the seventies and eighties (Twedt et al., 1977; Day, 1985) and framed the development of statistical seasonal hydrological forecasting. ESP is a very useful method to study the influence of meteorological boundary conditions, which are obtained from observed long-term records, on the results of the hydrological forecasting model. In contrast, the reversed-ESP

experiment is based on actual meteorological forcing but involves an ensemble of initial states, which makes it an appropriate method to study the influence of initial conditions on forecast results. The combination of both methods is also subject to recent research on predictability of hydrological systems (e.g. the VESPA approach, Wood et al., 2016)."

Page 2, line 18: I suggest adding "on the one hand", because otherwise the sentence might be misleading as seasonal forecast aren't solely an initial state problem (as you mention below).

Done.

Page 3, line 9: Please be more precise by adding "...based on seasonal predictions in the Alps" (or something similar).

Done.

Page 3, line 15: I suggest adding that the Inn basin belongs to the catchment of the Danube and that the Inn is the main tributary of the Upper Danube.

Good point. We added some additional text.

Page 4, line 5: Which "multi-year period" did you chose (the whole HISTALP period)?

We added "(i.e., 1996-2009)".

Page 4, line 6: Please add a reference and explain what you mean by "randomly selecting valid values".

We dropped this sentence.

Page 4, line 13-26: Please add the temporal and spatial resolution of the output from both GCMs you're using in your model.

The spatial resolution is already mentioned in the brief descriptions of the climate models. However, we will add a further remark that confirms the use of the original grid spacing for water balance simulations. Now both spatial and temporal resolution are described clearly.

"Monthly grids of the climate models with their original grid spacing (as specified above) are used as forcing data for the water balance model which is described in the next section."

Page 4, line 24: In line 28 you state, that only re-forecasts starting in November are considered, but here also "25 Oct." is listed as initial start date. Please explain.

The run initialised on 25 Oct is part of the lagged ensemble as the first (complete) month in the output is November. We will add the fact that a lagged ensemble is applied.

Page 4, line 31: What is the grid-size of the AWARE model used in this study? Page 5, 7-9: I suggest splitting this sentence in order to make it easier to read.

The fact that the spatial resolution of the AWARE model was missing in the discussion paper was also addressed by the comments of reviewer #2 We are sorry that this information was missing. The spatial resolution of the model setup for the Inn headwaters is 1000 m. The revised version of the manuscript will definitely include this information. Moreover, the sentence you have mentioned is split.

Page 5, line 27: Do you recognize mismatches in summer / autumn, when the reservoirs are filled-up, too? Please comment on this.

In summer and autumn, a mismatch is also expected due to reservoir operations. In contrast to spring, when precipitation is almost completely accumulated in form of snow and a recession line is obvious in the runoff hydrograph, a similar attribution is not possible in summer/autumn because the mismatch in reservoir operation might be obliterated by rainfall. In order to better address both effects explicitly, a better representation of reservoirs is required in the model. We already mentioned improving the model description of reservoirs in the outlook (p. 9, lines 29-30).

Page 6, line 17-20: I suggest adding the number of ensemble-members for each AWARE-run. This will help the reader remembering the set-up you described in section 2.2.3

The total number of members of each CM ensemble is now added to the list of runs.

Page 7, line 16: I don't see the information / added value of Fig. 3 (c) for the reader, because there's no comparison to measured SWE. Please explain why you decided to include this figure.

In the discussion paper, Fig. 3c) is referred when averaging of snow conditions in February are explained. We agree that this context does not necessarily require additional material (e.g., a figure) to support these explanations. However, the figure shows the variability of SWE in February. We now refer to the figure in this context.

Figure 3 (d): You should add the errors bars to the legend.

Done.

Page 8, line 8: As far as I understand, the hit rate for precipitation is equal (GloSea5) or lower (CFSv2) compared to temperature and higher as you state.

We are sorry for these misleading explanations. We rewrote the next sentence indicating that correlation is higher in case of precipitation only. Please refer to the next comment.

Page 8, line 8: As your finding that "The skill in precipitation predictions is higher [than temperature]" isn't something I would have expected before, you should extent discussing this result more detailed and refer to similar and contradicting results from other studies.

Regarding precipitation, it is well known that this is generally less skilfully predicted than temperature in most regions. We also acknowledge that while the results shown here indicate some skill, they do not show significantly different skill between temperature and rainfall and so this text has been altered:

“In the case of GloSea5-AWARE, the hitrate of correctly predicted anomalies regarding precipitation is […]”

Page 8, line 11-13: You state that predicting of hydrological storages (SWE in your case) is more robust / skillful than predicting fluxes; You should prove this statement. As your hydrological model also generates flows, I suggest relating this statement to your model results.

You are right! The paper would really benefit if we would provide a quantitative assessment which proves this statement. However, we think that it is not possible to simply compare the accuracy of storages to corresponding values of runoff. In winter runoff is subject to low flow conditions affected by reservoir operation. Focussing on other seasons would make this comparison more difficult due to the different climatological and hydrological processes. Thus, we suggest focussing on precipitation instead because we can compare the range of monthly scale correlations to the corresponding NDJF values. We will add the following lines:

“In our study, we found monthly scale correlations computed for precipitation forecasts ranging from -0.29 to 0.30 (GloSea5-AWARE) and -0.11 to 0.15 (CFSv2-AWARE), respectively. These are generally lower than the corresponding values achieved for the averaged NDJF forecasts (GloSea5-AWARE: 0.61, CFSv2-AWARE: 0.31). Similar values of the same order have been observed for SWE forecasts (GloSea5-AWARE: 0.57, CFSv2-AWARE: 0.28).”

Thank you for pointing us in this direction!

Page 8, line 26-28: Without doubt the representativeness of measured SWE and its interpolated on basin scale is problematic. But in my opinion you should mention, that the water balance model is never perfect and that it introduces uncertainties into hydrological forecasts, too. I guess it is your intention to exclude (at this state) the hydrological model-related errors by using a reference run?

We agree! Reviewer #2 also pointed us in this direction. At this stage, we exclude the hydrological model, which definitely introduces additional uncertainties. Instead, the comparison with the reference run (HISTALP-AWARE) is evaluated here. We rephrase this statement accordingly:

“However, the comparison between HISTALP-AWARE and the CM-based seasonal forecasts highlights GCM-forecast skill and acknowledges the fact that the water balance model is never perfect since it introduces uncertainties into hydrological forecasts, too. Due to the reasonably good agreement between seasonal forecasts and the reference run, the skill of CM-based forecasts is viewed promising.”

Page 8, line 28-29: I suggest to use “GCM-forecast skill” (or something similar) instead of “model skill when using CM-based forecasts”.

Done. Please also refer to the previous comment.

Page 8, line 31: The cumulative snow melt is very difficult to recognize in figure 5 as its value is very small. So why did you plot this parameter (you don't use this information in the text anymore)? I think it can be skipped.

You are right to say that snowmelt might be neglected in this chart. However, we think that this information is also important in this context since the water balance is closed. Otherwise the question about the relevance of snowmelt might arise.

Page 9, line 2: Please add "... and CFSv2-AWARE".

Done. This information was also added in other parts of the text in order to make clear that these values represent basin averages derived using AWARE.

Page 9, line 10: Please mention, which aspects of your method are really "new" (e.g. predictand, ...).

Indeed, the term "new" should be explained more detailed: We have added the following lines:

"SWE was chosen as predictand here and two independent climates model were used as input data for monthly scale distributed water balance model. A robust approach based on standardised anomalies was applied in order to bridge the gap in scale between GCMs and the water balance model."

Page 10, line 3: Interesting statement. Could you please comment on the definition of the "target accuracy". Who defined this Was is defined by users / stakeholders?

The target accuracy is not a strictly defined threshold. The value should reflect an improvement over a 50:50 probability. Moreover, 70% is a realistic value of forecasts skill in the mid-latitudes which can be provided by climate models (Bell et al., 2017).

References:

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, Hydrol. Earth Syst. Sci., 21, 4681–4691, doi:10.5194/hess-21-4681-2017, 2017.

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., Déqué, M., Domeisen, D. I. V., Fröhlich, K., Hendon,H., Imada, Y., Ishii, M., Iza, M., Karpechko, A. Y., Kumar, A., MacLachlan, C., Merryfield, W. J., Müller, W. A., O'Neill, A., Scaife, A. A., Scinocca, J., Sigmond, M., Stockdale, T. N., and Yasuda, T.: The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter?, Quart. J. Roy. Meteor. Soc., 142, 1413–1427, doi:10.1002/qj.2743, 2016.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Resour. Plann. Manage., 111, 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Domeisen D.I.V., Butler A.H., Fröhlich K., Bittner, M., Müller, W.A., Baehr, J.: Seasonal predictability over Europe arising from El Niño and stratospheric variability in the

MPI-ESM seasonal prediction system. Journal of Climate 28: 256–271, doi: 10.1175/JCLI-D-14-00207.1, 2015.

Förster, K., Hanzer, F., Stoll, E., Schöber, J., Scaife, A. A., MacLachlan, C., Huttenlau, M., Achleitner, S., Strasser, U.: Probabilistic retrospective forecasts of snow accumulation for the upcoming winter season in the Inn headwaters catchment (Austria). Geophysical Research Abstracts, 19, EGU2017-15495, 2017.

Greatbatch, R.J., Gollan, G., Jung, T., Kunz, T.: Factors influencing Northern Hemisphere winter mean atmospheric circulation anomalies during the period 1960/1961 to 2001/2002. Q. J. R. Meteorol. Soc. 138: 1970–1982, doi:10.1002/qj.1947, 2012.

Jörg-Hess, S., Griessinger, N., and Zappa, M.: Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas, J. Hydrometeorol., 16, 2169–2186, 2015.

Kang, D., Lee, M.I., Im, J., Kim, D., Kim, H.-M., Kang, H.-S., Schubert, S.D., Arribas, A.A., MacLachlan, C.: Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems. Geophysical Research Letters 10: 3577–3585, doi: 10.1002/2014GL060011, 2014.

Kim, H.-M., Webster, P. J., and Curry, J. A.: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, Climate Dyn., 39, 2957–2973, doi:10.1007/s00382-012-1364-6, 2012.

Kumar, A., Chen, M., Wang, W.: Understanding prediction skill of seasonal mean precipitation over the Tropics. J. Clim. 26: 5674–5681, 2013.

Molteni, F., Stockdale, T.N., Vitart, F.: Understanding and modelling extratropical teleconnections with the Indo-Pacific region during the northern winter. Clim. Dyn. 45: 3119–3140, doi: 10.1007/s00382-015-2528-y, 2015.

Riddle, E.E., Butler, A.H., Furtado, J.C., Cohen, J.L., Kumar, A.: CFSv2 ensemble prediction of the wintertime Arctic Oscillation. Climate Dynamics 41: 1099–1116, 2013.

Scaife A.A. et al.: Tropical Rainfall, Rossby Waves and Regional Winter Climate Predictions. Quart. J. Roy. Met. Soc., DOI: 10.1002/qj.2910, 2017.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519, doi:10.1002/2014GL059637, 2014.

Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H., Eade, R., Gordon, M., MacLachlan, C., Martin, N., Dunstone, N., and Smith, D.: Seasonal winter forecasts and the stratosphere, Atmos. Sci. Lett., 17, 51–56, doi:10.1002/asl.598, 2016.

Twedt, T. M., Schaake Jr, J. C., and Peck, E. L.: National Weather Service extended streamflow prediction, in: Proceedings Western Snow Conference, pp. 52–57, 1977.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, Journal of The Royal Society Interface, 11, 20131 162–20131 162, doi:10.1098/rsif.2013.1162, 2014.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, 15 L14 401, 5–5, doi:10.1029/2008GL034648, 2008.

Reply to anonymous reviewer #2

Reviewer's comments are in italics

This study explores the forecast skill of snow water equivalent (SWE) by using CGCM- driven water balance model simulations over a headwater region. While the topic is quite interesting and some results (e.g., GloSea5-driven forecasts) are potentially promising, the manuscript could be further improved after addressing several comments below.

We would like to thank Anonymous Reviewer # 2 for his/her detailed review of our manuscript. The comments will help us in the process of improving the discussion paper.

Major comments:
1. An interesting question that could be answered in this manuscript is whether precipitation or temperature prediction more important for the SWE forecasting over the headwater region. Although precipitation prediction is less skillful than temperature in many cases, the study region shows less skillful temperature prediction, perhaps due to the deficiency in snow or frozen soil processes. To compare their relative roles, precipitation or temperature forecast could be replaced with climatology before driving the SWE model. Such comparison would provide implications in advancing SWE forecasting.

We appreciate your suggestion! Reviewer #1 also asked us to address the relative roles of temperature and precipitation more explicitly. Indeed, a model experiment that includes both dynamical and climatological forcing data helps to analyse the relevance of temperature and precipitation for SWE forecasts. We will follow your suggestion and we performed such a model experiment. Accordingly, the following runs are performed:

1. temperature from models, precipitation from models (this is the configuration we have applied so far)
2. temperature from models, precipitation from climatology
3. temperature from climatology, precipitation from models

The second and the third model runs is analysed in the same way as already done with respect to the first model experiment. At this stage, we will also keep in mind your suggestions outlined in the second comment of your review. All performance and skill measures will be summarised in an additional table.

2. The study shows that GloSea5-driven SWE forecasting is better than the CFSv2- driven forecasting in terms of pearson correlation for the ensemble mean, but did not tell why the former is better? Some information on precipitation and temperature forecasts could be mentioned in the abstract. Moreover, probabilistic metrics (e.g., RPSS) is needed besides just simply using correlation. Given that this manuscript is not a short communication, I would encourage the authors to have a more comprehensive evaluation for SWE forecasting.

There are many differences between the CFSv2 and GloSea5 systems that could in principle explain the higher skill of the GloSea5 system and a full answer to this question is beyond the scope of our study, but one likely reason is that the skill of the NAO/AO is higher in Glosea5 (Scaife et al 2014) than in CFSv2 (Riddle et al 2013).

We also added more details on precipitation and temperature forecasts to the abstract:

"Even though predictions for precipitation may not be significantly more skilful than for temperature, the predictive skill achieved for precipitation is retained in subsequent water balance simulations when snow water equivalent (SWE) in February is considered."

We agree that there are many more skill measures which could be addressed in our analyses. We follow your suggestion to add some more metrics. As we are using the ensemble mean instead of using individual ensemble members, we decided to use the deterministic (single value) metrics where appropriate:

- The Continuous Ranked Probability Skill Score (CRPSS) is equivalent to the Mean Absolute Error (MAE) in case of a deterministic (single value) forecast, which is why CRPSS is used as measure representing the mean absolute error of forecasts. Here, we compute the MAE skill score (MAESS).
- The Ranked Probability Skill Score (RPSS) is equivalent to the Brier Skill Score (BSS) if a two categories forecast is considered. Thus, we will also compute the BSS values.
- We will also compute RMSE as suggested.

The revised version of the manuscript will include a table that provides these metrics. Similarly, the results of the model experiment suggested in your first comment will also analysed using these metrics. Thank you for these suggestions.

Minor comments:

3. Does the AWARE water balance model distinguish the input of liquid or solid precipitation? If so, how to obtain the solid precipitation from global climate forecast model like CFSv2?

Yes, it does. This information was still missing. We have added the method how the phase partitioning is performed in the model:

"For each grid cell the relative contributions of rainfall and snowfall are computed taking into account two threshold temperature values. If the air temperature falls below the lower threshold temperature, the monthly precipitation depth is assumed to be snowfall only. In contrast, air temperatures exceeding the upper threshold indicate rainfall only. In order to enable the occurrence of both snow and rain, a transition range between both thresholds is defined. Based on air temperature, the fraction of rain and snow is linearly interpolated between these both thresholds."

4. What is spatial resolution for the AWARE model over the study catchment?

We are sorry that is important information was also missing. Reviewer #1 also asked us to specify the spatial resolution of AWARE. It is 1000 m for the Inn headwaters.

5. What is the definition for the benchmark Nash-Sutcliffe efficiency?

In the revised version of the manuscript, we will provide the definition of the benchmark Nash Sutcliffe model efficiency in a new Appendix section along with the other metrics as suggested in comment #2.

6. For the benchmark NSE during the validation period, why does it drop to 0.25? Is it because there is trend or non-stationarity in the time series?

The benchmark Nash-Sutcliffe model efficiency is more sensitive to differences between two time series than the standard Nash-Sutcliffe model efficiency. We will explain the differences in the revised manuscript. The lower performance of the validation period might also be related to reservoir operations. We already addressed the need for a better reservoir representation in the outlook. Moreover, changes in glacier characteristics are not yet fully addressed by the water balance model. Since the calibration period was subjected to negative mass balances, positive mass balances have been observed in the 1980s. This refer to your suggestion to consider possible non-stationaries in the time series. We added the following lines to the manuscript:

Model description section:
"A possible reason for the lower E_b value might be the fact that the validation period has seen an advancing of glaciers due to positive glacier mass balances. In contrast, the calibration period is characterised by a shrinkage of glaciers volumes. Both processes are not incorporated in the model so far."

Outlook:
"Moreover, a better representation of changes in glaciated area is currently being investigated through coupling AWARE with a glacier evolution model developed by Marzeion et al. (2012)."

7. Figure 4. Besides correlation, how about the RMSE for the prediction?

Yes, we will add RMSE as well. Please refer to comment #2.

8. Figure 4. Is the model-simulated SWE or observed SWE used for verification? If the former, how to demonstrate the usefulness of the SWE forecasting given the limited skill in SWE simulation with AWARE (where NSE=0.25 in the validation period)?

HISTALP-AWARE computations of SWE were used to assess the skill. Reviewer #1 also commented on model uncertainty involved in hydrological modelling. We used the reference run (HISTALP-AWARE) for verification because further uncertainties are involved in running water balance model. The revised version of the manuscript will address model uncertainty more explicitly:

"However, the comparison between HISTALP-AWARE and the CM-based seasonal forecasts highlights GCM-forecast skill and acknowledges the fact that the water balance model is never perfect since it introduces uncertainties into hydrological forecasts, too."

Please also refer to our reply to your comment #6. We added some remarks regarding possible reasons that might explain the lower model performance.

References

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, Cryosphere, 6, 1295–1322, doi:10.5194/tc-6-1295-2012, 2012.

Riddle, E.E., Butler, A.H., Furtado, J.C., Cohen, J.L., Kumar, A.: CFSv2 ensemble prediction of the wintertime Arctic Oscillation. Climate Dynamics 41: 1099–1116, 2013.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519, doi:10.1002/2014GL059637, 2014.

Revised Version of the manuscript

Retrospective forecasts of the upcoming winter season snow accumulation in the Inn headwaters (European Alps)

Kristian Förster[1,2,3], Florian Hanzer[3,4], Elena Stoll[2], Adam A. Scaife[5,6], Craig MacLachlan[5], Johannes Schöber[7], Matthias Huttenlau[2], Stefan Achleitner[8], and Ulrich Strasser[3]

[1]Leibniz Universität Hannover, Institute of Hydrology and Water Resources Management, Hannover, Germany
[2]alpS - Centre for Climate Change Adaptation, Innsbruck, Austria
[3]Institute of Geography, University of Innsbruck, Innsbruck, Austria
[4]Wegener Center for Climate and Global Change, University of Graz, Austria
[5]Met Office Hadley Centre, Exeter, Devon, United Kingdom
[6]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom
[7]TIWAG, Tiroler Wasserkraft AG, Innsbruck, Austria
[8]Unit of Hydraulic Engineering, Institute of Infrastructure, University of Innsbruck, Innsbruck, Austria

Correspondence to: Kristian Förster (foerster@iww.uni-hannover.de)

Abstract. This article presents analyses of retrospective seasonal forecasts of snow accumulation. Re-forecasts with 4 months lead time from two coupled atmosphere-ocean general circulation models (NCEP CFSv2 and MetOffice GloSea5) drive the Alpine Water balance and Runoff Estimation model (AWARE) in order to predict mid-winter snow accumulation in the Inn headwaters. As the snowpack is a hydrological storage that evolves during the winter season, it is strongly dependent on precipitation totals of the previous months. Climate model (CM) predictions of precipitation totals integrated from November to February (NDJF) compare reasonably well with observations. This predictive skill Even though predictions for precipitation may not be significantly more skilful than for temperature, the predictive skill achieved for precipitation is retained in subsequent water balance simulations when snow water equivalent (SWE) in February is considered. Given the AWARE simulations driven by observed meteorological fields as a benchmark for SWE analyses, the correlation achieved using GloSea5-AWARE SWE predictions is $r = 0.57$. The tendency of SWE anomalies (i.e. the sign of anomalies) is correctly predicted in 11 of 13 years. For CFSv2, the corresponding values are $r = 0.28$ and 7 of 13 years. The results suggest that some seasonal predictions may be capable of predicting tendencies of hydrological model storages in parts of Europe.

1 Introduction

Seasonal Climate model (CM)-based seasonal predictions are an emerging new field in hydrology (e.g., Yuan et al., 2015; Svensson et al., 2015; Mackay et al., 2015) complementing current progress in predicting long-term developments in changed hydrological conditions as a consequence of anthropogenic greenhouse gas emission. In contrast to climate change projections, seasonal predictions focus on hydrological states of the upcoming months from their dependence on initial states (Warner, 2011). These can provide "Climate Services": A set of tools, products and information, serving decision makers and practitioners and bringing all types of information on climate research into practice at all levels of society (Vaughan and Dessai, 2014).

This makes them relevant for detecting anticipated short-term changes in hydrological systems as requested by international research programmes such as "World Climate Research Programme" (WCRP, e.g., Kirtman and Pirani, 2009), "Future Earth" (Greenslade and Berkhout, 2014), and more specialised programmes like, e.g., the International Network for Alpine Research Catchment Hydrology (INARCH, Pomeroy et al., 2015) which is part of the Global Energy and Water Cycle Exchanges Project (GEWEX, Chahine, 1992). In this context, seasonal predictions contribute, on the one hand, to cope with the WCRP "Grand Challenges" (Reid et al., 2010), and are used on the other hand to detect short-term changes in coupled hydrological-societal systems. The latter consideration of water and humans (Sivapalan et al., 2012) seeks to better understand interactions between society and hydrological systems for which seasonal predictions can also be seen as relevant. The goal of the current scientific decade "Panta rhei" of the International Association of Hydrological Sciences (IAHS) is to better understand these interactions on different scales of time (Montanari et al., 2013).

Seasonal outlooks of hydrological variables have been prepared for decades. Observed system states and rules with regional relevance have been utilised to prepare monthly outlooks (Pagano et al., 2004; Bell et al., 2017).A Antecedent hydrological and meteorological data have been used to predict monthly to seasonal streamflow using statistical methods (e.g. regression models) in various hydrological regimes (Pagano et al., 2004; Robertson and Wang, 2012; Schick et al., 2015). Another common way to predict future hydrological states is to run a process-based hydrological model based on known initial states and force it with ensembles of meteorological data observed in the past. This methodology is well known and referred to as Ensemble Streamflow Prediction (ESP, Wood and Lettenmaier, 2008). Its combination with meteorological data as forcing (reverse-ESP) The development of this method goes back to the seventies and eighties (Twedt et al., 1977; Day, 1985) and framed the development of statistical seasonal hydrological forecasting. ESP is a very useful method to study the influence of meteorological boundary conditions, which are obtained from observed long-term records, on the results of the hydrological forecasting model. In contrast, the reversed-ESP experiment is based on actual meteorological forcing but involves an ensemble of initial states, which makes it an appropriate method to study the influence of initial conditions on forecast results. The combination of both methods is also subject to recent research on predictability of hydrological systems (e.g. the VESPA approach, Wood et al., 2016). In the last decades, coupled Atmosphere-Ocean General Circulation Models (AOGCM) have become a viable method for seasonal predictions (Svensson et al., 2015). These climate model (CM)-based model-based forecasts provide future meteorological/climatological conditions for the next weeks (sub-seasonal forecasts), months (seasonal forecasts), or decades (decadal forecasts) on a physical basis rather than on statistics. An overview about the state-of-the-art of CM-based seasonal predictions is provided by Doblas-Reyes et al. (2013) and Yuan et al. (2015). Like numerical weather prediction, seasonal forecasts are on the one hand an initial-state problem since predictions of the atmospheric states of the upcoming months strongly depend on the initial states of the atmosphere, oceans, land, and sea ice. In contrast to weather predictions, the need for considering ocean and sea ice dynamics is crucial since these components of the climate system affect atmospheric phenomena on time scales beyond typical weather predictions. Another important difference from numerical weather predictions is the dependence of seasonal predictions on boundary conditions. Like long-term climate predictions which are based on anthropogenic greenhouse gas emissions, CM-based seasonal predictions require adequate definitions of boundary conditions (Doblas-Reyes et al., 2013).

The skill of CM-based seasonal predictions is not distributed equally in space and time (Smith et al., 2012; Kim et al., 2012; Kirtman et al., 2014). For instance, the skill in Europe is much lower than in the tropics, where phenomena like El Niño Southern Oscillation (ENSO) are predictable with higher accuracy (Yuan et al., 2015). Current progress on improving predictability has been recently reported by Scaife et al. (2014) who demonstrated skilful predictions of the North Atlantic Oscillation, a feature that is relevant for seasonal predictions in Europe. Bruno Soares and Dessai (2015) found that there is a mismatch in supply and demand regarding seasonal forecast products which is limited by skill levels in some regions although the authors also detected additional non-scientific reasons for this mismatch like, e.g., insufficient communication of forecasts to the users.

In the present study, we focus on a way to make seasonal predictions better exploitable for demands in context of hydro power generation in the Alps.For this application, it would be

In general, hydrological forecast models are quite sensitive to initial hydrological conditions such as antecedent rainfall, soil moisture and SWE. Uncertainties in the data of antecedent meteorological conditions influence the quality of process-based hydro-meteorological models in hourly resolution, e.g. in case of two days flood forecasts (Achleitner et al., 2012) or one month sub-seasonal streamflow drought forecasts (Fundel et al., 2013). Statistical seasonal streamflow forecast models can be improved when initial conditions with respect to soil moisture and groundwater flow (Robertson et al., 2013) or snow water equivalent (Pagano et al., 2004) are considered. Discharge from alpine catchments is known to be related to snow and ice melt (Viviroli et al., 2003; Kaser et al., 2010). For hydro power generation it is interesting to know if the upcoming winter season will be a winter season is above or below average regarding the accumulation of snow. The amount of water released as a consequence of snow melt in spring is a key parameter for the management of reservoirs. For water management demands such as efficient hydro power production, large efforts have been made to measure SWE in catchments of reservoirs (Painter et al., 2016; Krajči et al., 2017; Schattan et al., 2017), to simulate distributed SWE in basins of reservoirs and water in-takes (Schöber et al., 2012; Hanzer et al., 2016), to improve flood forecasts with distributed SWE data (Schöber et al., 2014) and to model future runoff under climate change conditions in snow and ice melt dominated catchments (Barnett et al., 2005; Schaefli et al., 2007; Finger et al., 2012; Hanzer et al., 2017). Gridded SWE data used for initialization of a process-based hydrological model improved predictions of SWE with lead times up to one month (Jörg-Hess et al., 2015). Seasonal streamflow and reservoir inflow predictions in snow-dominated basins were quite skillful during the snow-melt season and showed larger uncertainties during the rest of the year (Schick et al., 2015; Anghileri et al., 2016).

Besides hydro power, seasonal prediction of the accumulation of snow may be relevant to estimate the future evolution of snow depth on skiing slopes for the winter tourism business (Abegg et al., 2013; Marke et al., 2015). Well-focused, sustainable operation of technical snow production could be a means of significant savings with respect of energy costs and water use (Hanzer et al., 2014; Olefs et al., 2010).

In the present study, we focus on a way to make seasonal hydrological predictions better exploitable for demands in context of water resources planning in the Alps. We present a systematic evaluation in order to detect the predictability of above and below average snow accumulation which is expected to significantly influence runoff in spring and early summer. To achieve this goal, CM-based seasonal forecasts are employed as input data to a water balance model that predicts snow water equivalent (SWE) and runoff in the Inn headwaters. A new aspect of this work is the focus on *hydrological storages* hydrological storages instead of instantaneous hydrological fluxes . Considering hydrological storages together with seasonal predictions currently receives little attention (see, e.g., Sene et al., 2016; Mackay et al., 2015; Bell et al., 2017)and the seasonal prediction of SWE in general. It is expected that the focus on integrated storages (e.g., mid-winter snow accumulation) is more robust than considering instantaneous fluxes (e.g., precipitation, runoff) in seasonal predictions.

Moreover, we focus on the winter season as extratropical seasonal forecasts appear to have the highest skill in this season (e.g. Riddle et al., 2013; Scaife et al., 2014; Kang et al., 2014). There are a number of reasons for this, including winter being the season when the stratosphere is active, which is known to affect predictions (e.g. Domeisen et al., 2015; Scaife et al., 2016; Butler et al., 2016). The winter season also shows much stronger dynamical connections to the tropics, allowing high predictability of tropical rainfall (Kumar et al., 2013) to be transmitted into the extratropics (Greatbatch et al., 2012; Molteni et al., 2015; Scaife et al., 2017).

Based on that, the research question remains: Can we detect above or below average snow conditions on the basis of based on CM-based seasonal predictions in the Alps? To answer this, CM-based and hydrological modelling is applied in an alpine case study. In Section 2 the relevant information about the study area, the climate data, the CM-based seasonal predictions, the water balance model, and the methodology for detecting the predictability of snow accumulation are described. In Section 3, the results are presented, compiled and discussed. Finally, Section 4 provides concluding remarks and an outlook for future work.

2 Material and methods

2.1 Study area

The Inn headwaters catchment upstream of the Kirchbichl gauging station covers an area of 9,310 km^2 and is located in Switzerland and Austria (see, Fig. 1). The Inn river is the main tributary to the upper Danube. Elevations in the catchment range between 486 and 4049 m a.s.l., with a mean elevation of approximately 2000 m a.s.l. About 3% of the catchment area is covered by glaciers. During the winter season runoff is lowest since a major fraction of precipitation is accumulated to the snow cover. In spring snow melt causes an increase in runoff reaching its maximum in August when glacier melt is highest. For the period 1985–2009, the average areal precipitation and runoff amount to 1225 $\mathrm{mm\,yr^{-1}}$ and 1000 $\mathrm{mm\,yr^{-1}}$, respectively. In the second half of the 20th century, several reservoirs have been built in the study area. Their capacity in total is $638 \cdot 10^6\ \mathrm{m}^3$.

2.2 Climate data and seasonal predictions

2.2.1 Climate data

The climate data provided by the HISTALP project (Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region, Auer et al., 2007) is a suitable dataset for studies on climatology and long-term changes of temperature and precipitation in the Alps. The data has been compiled for a long period of time (1800-2010) and includes a dense observational station network from different countries in the Greater Alpine Region. Moreover, it has been quality checked and homogenised (Auer et al., 2007; Chimani et al., 2013). Mean temperature and precipitation depth are provided on a grid with a temporal resolution of one month and a spatial resolution of five minutes (approx. 6 km).

2.2.2 Climatological forecasts

In the framework of this study, the term "climatological forecasts" refers to simulations based on long-term averages of air temperature and precipitation depth for each month based on the HISTALP data. For instance, considering a climatological forecast for January, mean air temperature and precipitation depth are computed through averaging each variable over all Januaries in a multi-year period . This understanding differs from other definitions of climatological forecasts in which a climatological forecast is constructed by randomly selecting valid values for January in that period instead of averaging. (i.e. 1996–2009).

2.2.3 Climate model-based seasonal predictions

In this study, two different AOGCMs are utilised as input data for further analyses of seasonal predictions. As outlined earlier, the requirements of CM-based seasonal predictions exceed the extent of numerical weather predictions with respect to the forecast horizon and the number of subsystems of the climate system that need to be considered. Due to the extended forecast horizon, oceans and sea ice need to be incorporated in the models as well (see, e.g., Smith et al., 2012; Doblas-Reyes et al., 2013; Yuan et al., 2015). In this study, two different AOGCMs are applied independently:

- NCEP's (National Centers for Environmental Prediction) Coupled Forecast System model version 2 (CFSv2, Saha et al., 2014) is an operational seasonal prediction system. Forecasts are initialised four times a day. The horizontal resolution is 0.5°(approx. 40 km). In order to derive monthly forecasts, runs between the 8th day of the previous month and the 7th day of the current month are utilised in order to generate a lagged ensemble. This methodology is proposed by Yuan et al. (2013) who applied this method to re-forecasts. Since re-forecasts are only available for every 5th day, a typical ensemble of CFSv2 re-forecasts comprises 24 members per month. The archive of re-forecasts includes data from 1985 to 2009. The maximum lead time is 9 months.

- MetOffice Global Seasonal forecast system version 5 (GloSea5) is a seasonal prediction system that runs operationally at the MetOffice (MacLachlan et al., 2015; Scaife et al., 2014). Compared to CFSv2, it has a higher ocean horizontal resolution (0.25°, approx. 20 km). The data applied in this study was provided by the SPECS project ("Seasonal-to-decadal climate Prediction for the improvement of European Climate Services", http://www.specs-fp7.eu/) and covers the period between 1996 and 2010. Re-forecasts for winter were used with initial start dates: 25 Oct., 01 Nov., and 09 Nov. For each date, 3 runs are available which gives an a lagged ensemble of 9 members per winter. This subset of hindcasts has a lead time of 4 months for each run.

Systematic analyses are performed for 1996 – 2009 (the period in which both models are available). Only those re-forecasts that start in November are considered. The lead time is limited to 4 months to predict snow conditions in February. Monthly grids of the climate models with their original grid spacing (as specified above) are used as forcing data for the water balance model which is described in the next section.

2.3 Water balance simulations using AWARE

The Alpine Water balance and Runoff Estimation model (AWARE, Förster et al., 2016) is a deterministic hydrological model operating on a regular grid at one month time steps. The model has been designed to estimate anomalies in hydrological variables at the catchment scale from anomalies in meteorological fields predicted by climate models. The coarse temporal resolution allows one to carry out seasonal predictions considering a large number of individual runs at a minimum of computational costs which justifies the coarse time step. As the study's focus is on anomalies in seasonal characteristics, using a monthly scale water balance model is feasible (Kling et al., 2012; Bock et al., 2016) and these models are also applied for seasonal hydrological predictions (Bell et al., 2017).

Required meteorological forcing data include both mean monthly air temperature and monthly precipitation totals provided as grids or station data, which makes the model parsimonious with respect to data requirements. Altitudinal gradients are applied in order to realistically redistribute temperature and distribution precipitation on the model gridresulting . In general, this feature results in a decrease in temperature with increasing elevation and an increase in precipitation on the mountains. For each grid cell the relative contributions of rainfall and snowfall are computed taking into account two threshold temperature values. If the air temperature falls below the lower threshold temperature, the monthly precipitation depth is assumed to be snowfall only. In contrast, air temperatures exceeding the upper threshold indicate rainfall only. In order to enable the occurrence of both snow and rain, a transition range between both thresholds is defined. Based on air temperature, the fraction of rain and snow is linearly interpolated between these two thresholds. Even though the model is also capable of reading shortwave radiation fields (Förster et al., 2016) in order to improve ice melt predictions, a simplified snow- and ice-melt simulation using air temperature only is possible. This simplification considers the fact that air temperature and precipitation are readily available and more predictable compared to some other meteorological fields. In order to perform simulations with this minimal input of data, the Thornthwaite (1948) evapotranspiration approach is applied. The soil water balance is computed following the approach of McCabe and Markstrom (2007). A linear storage is applied in order to account for the recession of runoff typically related to groundwater processes.

The model spatial resolution of the Inn headwaters setup in the AWARE model is 1000 m. Besides a grid based model domain, AWARE assumes a baseline (reference) meteorological dataset for calibration, which is shown in Figure 2 using the HISTALP data from 1996 to 2009 as the reference period (this run is herein referred to as HISTALP-AWARE). The Nash-Sutcliffe model efficiency (NSE) amounts to $E = 0.92$ which might be viewed as very good model performance. As suggested by Schaefli and Gupta (2007), the benchmark Nash-Sutcliffe model efficiency is computed as well ($E_b = 0.45$). This benchmark NSE value accounts for strong effects of seasonality . (Eq. A1 in the Appendix Sect. A). While the standard NSE indicates if a model is better than the average of observed values, the benchmark NSE proves if the model performance exceeds the corresponding value of a simple model that simply predicts long-term averages for each month. Since the benchmark NSE is also greater than zero, the model is more skillful than applying long-term averages. According to Klemeš (1986) a split sample test is applied including an independent validation period ranging from 1984 to 1995. The corresponding NSE and benchmark NSE are 0.91 and 0.25$E = 0.91$ and $E_b = 0.25$, respectively. As the model A possible reason for the lower E_b value might be the fact that the validation period has seen an advancing of glaciers due to positive glacier mass balances. In contrast, the calibration period is characterised by a shrinkage of glaciers volumes. Both processes are not incorporated in the model so far. However, as the model performance of the validation period is still comparable to the calibration period, the model is found to be suitable for predictions. The mismatch of runoff simulations in winter, especially in March can be attributed to the effects of reservoirs on river flow in the catchment area which are not represented in the model so far. In this period water is released from seasonal storages which are filled in summer.

Another advantage of the one month time step is the lower complexity with respect to downscaling of climate model data. Current approaches focus on statistical (e.g., Crochemore et al., 2016) or dynamical downscaling (e.g., Förster et al., 2014) of coarse atmospheric data fields (e.g., derived by climate models). AWARE builds upon a simple and robust approach which is based on anomalies. For instance, Marzeion et al. (2012) successfully add anomalies from other datasets to a reference climatology to compute glacier mass balances at the global scale. In order to account for different spreads of distributions, standardised anomalies are considered in our study. According to Wilks (2006) this approach is feasible when "working simultaneously with batches of data that are related, but not strictly comparable". This is a typical situation in case of observational data and re-forecasts. Standardised anomalies z_x are simply computed for a variable x, taking into consideration its long-term mean $\bar{x}$ for a given month and the corresponding empirical standard deviation $\tilde{s}_x$ (Wilks, 2006):

$$z_x = \frac{x - \bar{x}}{\tilde{s}_x} \tag{1}$$

Given that two datasets x and y are comparable (e.g., reference climatology and the climatology of re-forecasts), their standardised anomalies z_x and z_y might be viewed comparable as well. Based on the assumption that $z_x = z_y$, Eq. 1 can be rearranged to

$$x = \frac{y - \bar{y}}{\tilde{s}_y} \cdot \tilde{s}_x + \bar{x}. \tag{2}$$

Anomalies of the climate model (i.e. $y - \bar{y}$) can easily be transformed to the climatology of the reference data set (i.e., x). Mean values and standard deviations are computed separately for each month and climate data set including HISTALP, CFSv2, and GloSea5. In this way, anomalies predicted by the climate models can be reliably transformed to typical anomalies of the observational data.

2.4 Model experiment for analysing the predictability of snow accumulation

The long-term simulations of the water balance provide monthly snapshots of valid system states for each state variable at any point in time. For each CM-based seasonal prediction run starting in November, system states for SWE, soil moisture, and groundwater storage computed for October are defined as initial states. In total four AWARE runs driven with different forcing datasets are available for each winter season between 1997 and 2009 (November to February, NDJF):

1. HISTALP-AWARE: Long-term continuous run based on observed HISTALP data (see, Sect. 2.2.1 and Fig. 2),
2. CF-AWARE: Climatological forecasts (CF) with average conditions computed using HISTALP (see, Sect. 2.2.2),
3. GloSea5-AWARE: CM-based seasonal forecast using GloSea5 (ensemble mean of 9 members), and
4. CFSv2-AWARE: CM-based seasonal forecast using CFSv2 (ensemble mean of 24 members).

The ensemble provided by each CM-based seasonal forecasts forecast of meteorological quantities is averaged prior to the water balance simulations. In general, ensemble seasonal predictions are subject to low signal to noise ratios. The signal in the ensemble mean is small in most cases and using members individually will mask the signal (Scaife et al., 2014; Eade et al., 2014). In general, each ensemble member of input data is individually processed in hydrological forecasting, which is why the averaging is typically implemented afterwards. However, a skill improvement is reported by recent seasonal prediction studies (e.g., Bell et al., 2017) in which the concept of averaging is applied prior to hydrological simulations. This approach seems feasible given that the time step of hydrological simulations is one month. Although the hydrological model is a conceptual model that mimics the basic physical principles, the temporal scale does not allow to capture the full dynamics of hydrological processes that are typically studied on smaller scales. Thus, the coarse temporal resolution of the modelling approach is to a certain degree “statistical” in nature which justifies the application of mean ensemble inputs. Moreover, the utilisation of standardised anomalies applied to CM-based seasonal forecasts in the AWARE model accounts for variance corrections to the ensemble mean values as suggested by Eade et al. (2014). Appropriate uncertainty can also be added to the predictions to ensure reliable probabilistic forecasts.

The basin-average time series of these water balance simulations are directly comparable. While the continuous long-term simulation represents a reference run (#1) serving as benchmark for seasonal predictions, the climatological forecasts (#2) help to judge whether anomalies will be above or below average. Correlations between the reference run and the water balance simulations forced by CM-based forecasts (#3 and #4) are computed to assess the predictive skill. Moreover, the tendency or sign of anomalies is compared through counting the coincidence of above (below) average anomalies in the reference run and the seasonal predictions.

A set of skill measures is used throughout the study in order to quantify the model skill of the different forecasts (CF-AWARE, GloSea5-AWARE, CFSv2-AWARE). Besides correlation and hitrate (i.e. the number of correctly predicted states divided by the total number of winters) other measures to assess the skill of the models are considered. For instance, the standard deviation of a single time series is a measure to compare the variability of forecasts. In contrast, the Root Mean Square Error (RMSE) also involves observed time series and gains insight into the absolute difference between time series. Since quadratic differences are summarised, a greater weight is assigned to larger differences, thus making RMSE sensitive to greater mismatches. In order to show the accuracy of the models for predicting the tendencies of anomalies (hitrate), the Brier Skill Score (BSS) is also computed (see Eq. A2 and Eq. A4 in Appendix Sect. A along with a brief description). In general, a skill score judges the improvement of a forecast system relative to a reference (climatology). A value of zero would indicate that the forecast system is not better than the reference. In contrast, a value of one indicates a perfect match of forecasts and observations. The BSS is related to the hitrate which has already been defined (higher hitrates go in hand with higher BSS). Finally, the Mean Absolute Error Skill (MAE, Eq. A3) is comparable to RMSE but does not account for quadratic weighting of differences. Like BSS, it can be computed as skill score MAESS (Eq. A4) that is a measure for the differences in absolute terms. In this way, it is less sensitive than RMSE to large differences but rather includes a reference run.

3 Results and Discussion

3.1 Long-term simulations and climatological forecast of SWE

While the applicability of AWARE to reconstruct the water balance in terms of observed runoff time series was demonstrated in Sect. 2.3, it is necessary to evaluate the model experiments HISTALP-AWARE and CF-AWARE with respect to SWE prior to the analyses of CM-based SWE forecasts. Figure 3(a) demonstrates the annual cycle of modelled SWE. The black dashed line is the mean value of all years computed using the reference run (HISTALP-AWARE). It compares well with the black bold line which represents the climatological simulations based on AWARE using average air temperature and precipitation depth for each month (CF-AWARE). Thus, the a climatological forecast is suitable to compute average snow conditions(see also Fig. . Figure 3(c) which includes shows the spatial distribution of average SWE in February). . The averages of SWE on the model highlights the typical snow distribution with highest values on the mountains and lower values in the valleys. Full time series are shown in Fig. 3(b). The boundary conditions of the climatological forecast are equal in each year. However, the initial conditions differ according to the initialisation each year in October which is obtained from the long-term run. Figure 3(d) depicts a subset of SWE observations compiled by Schöber et al. (2016). In contrast to the cited study, which explains the methodology of SWE sampling in detail, here only stations above 1400 m a.s.l. have been selected in order to better match the average catchment elevation (Sect. 2.1). The correlation between computed SWE in February and the SWE observations in February is $r = 0.65$ (Fig. 3(b) vs. Fig. 3(d)). This comparison should be interpreted with caution. First, despite the fact that a sub-selection of stations that better match the mean elevation of the catchment has been chosen for this analysis, the full range of elevation bands in the basin is not fully covered by the observational dataset. Moreover, scaling issues limit spatial and temporal representativeness, since averaged point-scale measurements recorded on a weekly scale are compared to basin-scale water balance simulation with one month time step. However, observed and computed SWE compare reasonably well. This underlines the applicability of AWARE to predict SWE.

3.2 CM-based seasonal predictions using AWARE

In a next step, anomalies computed using AWARE forced by CM-based seasonal forecasts are compared to the corresponding values of the reference run (HISTALP-AWARE, #1). This evaluation is demonstrated in Fig. 4 for temperature, precipitation depth, and SWE in February. Anomalies in temperature and precipitation depth refer to the period November to February (NDJF) in each winter and represent average values at the basin scale (i.e. the mean of all grid points of the meteorological fields in AWARE). In this way, the values are subject to the statical transformations and elevation dependent redistributions as outlined in Section 2.3. The anomalies of the reference AWARE run driven by HISTALP are shown in the top panels of Fig. 4 (HISTALP-AWARE). Their correlation is set to 1 by definition since this run is viewed as reference. The seasonal forecasts computed using AWARE driven by GloSea5 GloSea5-AWARE (center) and CFSv2 CFSv2-AWARE (bottom) are also displayed. In addition, Tab. 1 (first model experiment column) provides a summary of skill measures for temperature, precipitation, and SWE.

Correlation coefficients computed for NDJF temperature anomalies range from $r = 0.17$ (CFSv2CFSv2-AWARE) to $r = 0.32$ (GloSea5GloSea5-AWARE). Tendencies in anomalies (i.e. the prediction of correct signs of anomalies) also vary between the models. This becomes obvious when counting the shaded areas indicating a mismatch between the seasonal forecast and the reference run. While GloSea5 GloSea5-AWARE correctly predicted the sign of temperature anomalies in 9 of 13 winters, the hitrate (i.e. the number of correctly predicted states divided by the total number of winters) achieved for CFSv2 achieved for CFSv2-AWARE only amounts to 8 of 13.

The skill in precipitation predictions is higher regarding both correlation and hitrate. 13 (see, Tab. 1). The differences between GloSea5-AWARE and CFSv2-AWARE in terms of standard deviation are small. Hence, both model settings show a similar variability of forecasts which can be attributed to the standardised anomaly approach. GloSea5-AWARE shows a smaller RMSE than CFSv2-AWARE does. A similar ranking of skill is obvious when considering the skill scores BSS and MAESS. The latter suggests that both model runs (GloSea5-AWARE and CFSv2-AWARE) are less skilful than climatology (MAESS < 0). However, the positive BSS values highlight the capability of predicting the tendency of temperature anomalies.

In the case of GloSea5GloSea5-AWARE, the hitrate of correctly predicted anomalies regarding precipitation is 9 of 13 ($r = 0.61$). As for temperature, the model skill of precipitation predictions computed by CFSv2 CFSv2-AWARE is also lower (hitrate 7 of 13, $r = 0.31$). This finding holds also true for the other skill measures, namely RMSE, BSS, and MAESS. However, the number of correctly predicted tendencies achieved using GloSea5-AWARE might be viewed as good results since the seasonal forecasts includes a lead time of 4 months. Single months show lower scores, suggesting that a temporal integration improves the robustness of results consistent with our approach using hydrological storages rather than fluxes. In our study, we found monthly correlations computed for precipitation forecasts ranging from -0.29 to 0.30 (GloSea5-AWARE) and -0.11 to 0.15 (CFSv2-AWARE), respectively. These are generally lower than the corresponding values achieved for the averaged NDJF forecasts (GloSea5-AWARE: 0.61, CFSv2-AWARE: 0.31). Similar values of the same order have been observed for SWE forecasts (GloSea5-AWARE: 0.57, CFSv2-AWARE: 0.28).

Given the skill measures from Tab. 1 (first column) and the coincidence of anomalies highlighted in Figure 4(c) highlights that the predictive skill achieved for precipitation depth is also prevailing for SWE in February. Even though correlation coefficients are slightly lower compared to precipitation depth (GloSea5GloSea5-AWARE: $r = 0.57$, CFSv2CFSv2-AWARE: $r = 0.28$), SWE values in February computed by AWARE driven by CM-based forecasts compare well to those of the reference run (HISTALP-AWARE). The hitrate achieved using GloSea GloSea-AWARE even reaches 11 of 13 while the hitrate of CFSv2 CFSv2-AWARE remains at the level of 7 of 13. An increase in skill in terms of RMSE, BSS, and MAESS is also obvious - at least partially

- for both models indicating that some skill measures suggest that SWE predictions are more robust than precipitation predictions.

A Bernoulli experiment helps to judge whether these hitrates differ from the performance of a “fair coin” for predicting above and below average conditions. The null hypothesis is: The hitrate of the seasonal forecasts does not differ from a random 50:50 probability (binomial test). Given the total number of winters $n = 13$ and a level of significance of $\alpha = 0.05$, the null hypothesis is rejected for hitrates above 9 of 13. This means that according to the results shown in Fig. 4 and Tab. 1, for seasonal predictions of SWE using GloSea5 this test rejects the null hypothesis, indicating significant skill. In contrast, the scores for CFSv2 are not significant.

Regardless the limitations discussed with respect to observed SWE, the correlations are much lower if the observations from Fig. 3(d) are involved in skill computations. The correlation between observed anomalies and GloSea5 is $r = 0.21$ while the corresponding value achieved using CFSv2 is only $r = 0.11$. These values are much lower than the correlations achieved using the reference run (HISTALP-AWARE). This finding might also be related to possible mismatches in representativeness between observations and simulations. However, the comparison between HISTALP-AWARE and the CM-based seasonal forecasts highlights model skill when using CM-based forecasts. This comparison revealed a GCM-forecast skill and acknowledges the fact that the water balance model is never perfect since it introduces uncertainties into hydrological forecasts, too. Due to the reasonably good agreement between seasonal forecasts and the reference runand , the skill of CM-based forecasts is viewed promising.

Finally, Fig.Figure 5 depicts time series of the water balance of the snow storage for each year and each AWARE model run. Monthly precipitation (separated for divided into rainfall and snowfall), cumulative snowmelt, and SWE are plotted. Moreover, the snow accumulation of the reference run (HISTALP-AWARE, #1) and the climatological forecast (CF-AWARE, #2) are displayed. The latter is subject to the same forcing in each year but is initialised according to the system states of AWARE in late fall. If the SWE computed by HISTALP-AWARE exceeds the corresponding value of CF-AWARE, above average snow accumulation prevails. Accordingly, the opposite is true for below average conditions. A similar comparison is possible for the predictions of GloSea5-AWARE and CFSv2CFSv2-AWARE. If the CM-based forecast and HISTALP-AWARE simultaneously indicate either above or below average conditions, a label “HIT” is added to the corresponding seasonal forecast. The overall hitrate is readable from FigTab. 41. Even though monthly precipitation depth differs between HISTALP-AWARE and the CM-based forecasts, the NDJF precipitation totals might compensate this monthly scale differences resulting in a good match of SWE values in February. This is obvious for many of the winter seasons shown in Fig. 5 (e.g., 1998/99 and 2000/01) and confirms the previous finding that improved model skill is possible when storages instead of instantaneous fluxes are considered.

3.3 The role of temperature and precipitation for SWE forecasts

In order to show the importance of both temperature and precipitation in SWE forecasting, Tab. 1 summarises the skill measure previously introduced for two other model experiments in which either temperature or precipitation is replaced by climatological forecasts: (i) Temperature from climatology is combined with precipitation forecasts from the climate models (second column of Tab. 1) and (ii) Precipitation from climatology is combined with temperature forecasts from the climate models (third column of Tab. 1). If one variable is replaced by climatology the standard deviation of anomalies is zero since the climatological forecasts have no deviations from climatology. This is in line with zero skill in terms of BSS and MAESS (see, temperature skills in the second column and precipitation skills in the third column, respectively). The skill measures of the respective variable that has not changed in this way is subject to the same characteristics as the full dynamical run (first column). For instance, if temperature is replaced by climatology, precipitation skills are equal to the full dynamic run (e.g., compare column one and two for precipitation).

In case of SWE, the effects of replacing either temperature or precipitation differ in terms of model skill. First, a drop in correlation is obvious in both cases. If temperature is replaced by climatology the hitrate of GloSea5-AWARE decreases only slightly to 10 but remains 7 in case of CFSv2-AWARE. If precipitation is replaced by climatology hitrates decrease in both cases and the standard deviation is much lower than in the full dynamic run. This indicates that the variability in SWE forecasts is mainly prescribed by precipitation in the current study setup. However, the influence of temperature would likely increase for predictions of SWE in the ablation season.

3.4 Model skill and its relation to other studies

Compared to findings reported in the literature, the results achieved in this study are promising given that the skill in Europe is generally found to be low. For instance, according to Weisheimer and Palmer (2014) the skill of DJF temperature is “marginally useful” using ECMWF’s System4. The rating for DJF precipitation is even found to be “not useful” (cf., Fig. 5 in Weisheimer and Palmer, 2014). Similarly, Kim et al. (2012) found some skill in terms of correlation for wintertime temperature predictions using System4. However, their study also suggests low absolute correlation coefficients for precipitation forecasts and for both temperature and precipitation forecasts achieved using CFSv2. A direct comparison to the results presented in this study is not possible since GloSea5 was not addressed in these studies. Moreover, given that only one single catchment is considered, a ranking of models is beyond the scope of this article. The predictability for SWE detected in this study can be related to both some skill in precipitation prediction and previous findings found for the persistence in SWE predictions for smaller forecast horizons. For instance, in case of the alpine snow cover, Jörg-Hess et al. (2015) underline the persistence in SWE predictions at least up to a lag of two weeks.

4 Conclusions

In this study, a systematic evaluation of CM-seasonal CM-based seasonal winter forecasts starting in November has been performed using a water balance model. A new method has been developed focussing on hydrological storages instead of instantaneous hydrological fluxes. SWE was chosen as predictand here and two independent climate models were used as input data for monthly scale distributed water balance model. A robust approach based on standardised anomalies was applied in order to bridge the gap in scale between GCMs and the water balance model. In this way, basin-scale averages of temperature and precipitation depth are temporally integrated in order to achieve November to February (NDJF) averages and totals, respectively. Given a lead time of 4 months, the application of the water balance model then allows predicting SWE in February which is relevant for many sectors like water management or hydro power generation. Based on year-by-year evaluation of re-forecasts using correlation analyses different skill measures and a Binomial test, the results achieved using GloSea5 and CFSv2 GloSea5-AWARE and CFSv2-AWARE indicate that dynamical (CM-based) seasonal forecasts can provide skill. A sensitivity analysis using different configurations of input datasets showed that SWE forecasts benefit from the skill in precipitation forecasts, especially in terms of variability and hitrate / Brier Score. These findings might be related to the hydro-climatological characteristics of the study area where snow accumulation is the major process during winter while snowmelt as a strong temperature dependent process is less important in this time (Fig. 5). In other environments the relative role of temperature and precipitation might look different.

Regarding predictability, the location of the study area is also of particular interest in the process of interpreting the results. Due to the fact that the Alps are situated in a transition zone between northern and southern Europe, the influence of large-scale climate patterns, such as the North Atlantic Oscillation (NAO), should be analysed more detailed in the future. It is also known that El Niño Southern Oscillation (ENSO) impacts the climate in Europe in late winter and stratospheric sudden warmings are also important (Ineson and Scaife, 2008; Scaife et al., 2016). A first assessment of possible connections between the NAO on the one hand and snow and glacier related states on the other hand only resulted in low correlations (c.f. Beniston and Jungo, 2002; Scherrer et al., 2004; Bartolini et al., 2009; Marzeion and Nesje, 2012). However, in the southern and western parts of the Alps this relationship between NAO and snow and ice properties might be explained more clearly. Recent skill improvements regarding CM-based seasonal predictions might explain our detectable skill (Scaife et al., 2014). Future work should address climatological processes that are related to model skill and involve other basins in different parts of the Alps.

Besides studying the climatological perspective of predictability, the results also revealed uncertainties involved in hydrological modelling using the water balance model and scaling issues regarding the representativeness of point scale SWE observations. These findings also suggest improvements regarding both the provision of basin-scale SWE observations and the water balance model as an outlook for future work. Low flow conditions in March might be better predicted if the model would account for artificial reservoirs in the study area. This feature Moreover, a better representation of changes in glaciated area is currently being investigated through coupling AWARE with a glacier evolution model developed by Marzeion et al. (2012). These features will be added to the model in the future.

However, the results of this study show that it is possible to detect skilful signals from dynamical (CM-based) seasonal predictions of hydrological storages in Europe, where seasonal predictions are still challenging. The results suggest that some seasonal predictions may be capable of predicting tendencies of hydrological model storages, although the skill of these predictions is in many cases low in Europe. The basic idea of this study is that a focus on hydrological storages rather than on hydrological fluxes might help in exploiting seasonal predictions. The first results of the methodology are promising with respect to practical applications in which hitrates above 70% might be seen as a reasonable target accuracy. Since snowmelt predictions are of particular interest in the study area, a similar approach could be applied to CM-based seasonal forecasts initialised in May. Future research should also address predictability studies in other regions. Moreover, it would be interesting to study the predictability of other hydrological storages like, e.g., glaciers, lakes or groundwater. A focus on probabilistic forecasts also is an interesting prospect for the future.

5 Data availability

- CFSv2 re-forecast data: https://nomads.ncdc.noaa.gov/data/cfsr-rfl-mmda/flxf/
- HISTALP: http://www.zamg.ac.at/histalp/index.php

Appendix A: Model performance and skill measures

The definition of the Nash-Sutcliffe model efficiency E and the benchmark Nash-Sutcliffe model efficiency E_b reads (Schaefli and Gupta, 2007):

$$E_b = 1 - \frac{\sum_{t=1}^{N} \left[q_{obs}(t) - q_{sim}(t)\right]^2}{\sum_{t=1}^{N} \left[q_{obs}(t) - q_{bench}(t)\right]^2} \tag{A1}$$

In this computation time series of observed q_{obs} and modelled q_{sim} quantities are considered for all time steps t. $q_{\text{bench}}(t)$ is the time dependent benchmark value at timestep t. $q_{\text{bench}}(t)$ is a long-term average computed for the month of time step t. The original definition of Schaefli and Gupta (2007) refer to daily series for which the long-term average for a specific calendar day is applied. According to Schaefli and Gupta (2007) E_b indicates if the model “has greater explanatory power than already contained in the seasonality of the driving forces (the climate)”. If $q_{\text{bench}}(t) = \bar{q}_{\text{obs}}$ is assumed, E_b is equal to the Nash-Sutcliffe model efficiency E. In contrast to E, E_b presumes the climatological mean of each time step as a benchmark against which all elements of the time series are compared. Since seasonality is inherent in many time series, generally $E_b \leq E$ holds.

A widely used measure to evaluate the accuracy of forecasts is the Brier Score B (Wilks, 2006):

$$B = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2 \tag{A2}$$

It is a special case of the Ranked Probability Score (RPS, see, e.g., Hersbach, 2000) that restricts the evaluation of forecasts to two categories (e.g., above or below average). The forecast f_i computed for each year i is compared to the corresponding state observed in that year o_i, whereby f_i and o_i are dichotomous states (0 or 1). The Brier score B is the average of the squared differences between f_i and o_i. The average refers to a range of N years. The best value that can be achieved in this way is zero indicating a perfect forecast skill. In contrast, 1 would indicate that all forecasts are wrong.

Another skill measure for forecasts is the Mean Absolute Error (MAE) which characterises, similar to RMSE, differences between the forecasted value q_f and the observed value q_o (in units of the underlying time series):

$$\mathsf{M} = \frac{1}{\mathsf{N}}\sum_{\mathsf{i}=1}^{\mathsf{N}} |\mathsf{q}_{\mathsf{f},\mathsf{i}}(\mathsf{t}) - \mathsf{q}_{\mathsf{o},\mathsf{i}}(\mathsf{t})| \tag{A3}$$

If $|q_{f,i}(t) - q_{o,i}(t)|$ is replaced by $(q_{f,i}(t) - q_{o,i}(t))^2$ and the square root is calculated from Eq. A3, this equation yields the Root Mean Square Error (RMSE). In contrast to the latter, MAE is less sensitive to larger differences between $q_{f,i}$ and $q_{o,i}$. Moreover, the MAE is comparable to the Continuous Ranked Probability Score (CRPS) used for probabilistic forecasts (Hersbach, 2000; Trinh et al., 2013) and can be used for single-value (deterministic) forecasts.

In order to compare these skill measures computed for different forecasts to a reference forecast (i.e., climatology), a skill score S measure is typically calculated. For instance, the MAE skill score (S_{MAESS}) can be derived using

$$\mathsf{S}_{\mathrm{MAESS}} = 1 - \frac{\mathsf{M}_{\mathrm{forecast}}}{\mathsf{M}_{\mathrm{reference}}}, \tag{A4}$$

with M_{forecast} is the MAE of the forecast system and $M_{\mathsf{reference}}$ is the MAE of the climatological forecast. Similarly, Eq. A4 can be applied to derive a Brier Skill Score S_{BSS} through replacing M by B.

Author contributions. K. Förster prepared the manuscript with contributions from all co-authors, designed the study, performed the water balance simulations and predictability analyses. He and F. Hanzer developed the AWARE model which has been designed for this kind of study. E. Stoll contributed to downscaling of climate model output and reviewed the literature with respect to connections between snow and glacier observations on the one hand and the NAO index on the other hand. A. A. Scaife and C. MacLachlan computed and provided the GloSea5 re-forecasts and helped with data usage, interpretation of the results and improving the methodology. Snow observations in the study area were evaluated by J. Schöber who also contributed to interpreting anomalies in SWE. M. Huttenlau coordinated the project. S. Achleitner and U. Strasser are key researchers of the project. They supervised the scientific work and helped discussing the results and improving the methodology.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was carried out as part of the project "W01 MUSICALS II – Multiscale Snow/Ice Melt Discharge Simulation for Alpine Reservoirs" at alpS – Centre for Climate Change Adaptation in Innsbruck, Austria. The K1-Centre alpS is funded through the Federal Ministry of Transport, Innovation and Technology (BMVIT), the Federal Ministry of Science, Research and Economy (BMWFW), as well as the Austrian federal states of Tyrol and Vorarlberg within the scope of COMET – Competence Centers for Excellent Technologies. The Programme COMET is managed by the Austrian Research Promotion Agency (FFG). We want to thank TIWAG – Tiroler Wasserkraft AG for the collaboration and co-funding the project. Another thanks goes to the NOAA (National Oceanic and Atmospheric Administration)

National Centers for Environmental Prediction (NCEP) for the provision the CFSv2 data. The retrospective forecasts of the GloSea5 model were kindly provided by SPECS project ("Seasonal-to-decadal climate Prediction for the improvement of European Climate Services", http://www.specs-fp7.eu/). We would like to thank Felix Oesterle who wrote the script to automatically retrieve and convert CFSv2 data. Assistance with HISTALP data provided by Anna-Maria Tilg and Barbara Chimani is greatly appreciated. A.A.S. and C.M. were supported by the joint DECC/Defra MetOffice Hadley Centre Programme (GA01101). The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover. We wish to thank two anonymous reviewers for their helpful comments that helped to improve the manuscript.

References

Abegg, B., Steiger, R., and Walser, R.: Herausforderung Klimawandel: Chancen und Risiken für den Tourismus in Graubünden, Amt für Wirtschaft und Tourismus, 2013.

Achleitner, S., Schöber, J., Rinderer, M., Leonhardt, G., Schöberl, F., Kirnbauer, R., and Schönlaub, H.: Analyzing the operational performance of the hydrological models in an alpine flood forecasting system, J. Hydrol., 412-413, 90–100, doi:10.1016/j.jhydrol.2011.07.047, 2012.

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., and Lettenmaier, D. P.: Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, Water Resour. Res., 52, 4209–4225, doi:10.1002/2015WR017864, 2016.

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, Int. J. Climatol., 27, 17–46, doi:10.1002/joc.1377, 2007.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309, doi:10.1038/nature04141, 2005.

Bartolini, E., Claps, P., and D'Odorico, P.: Interannual variability of winter precipitation in the European Alps: relations with the North Atlantic Oscillation., Hydrol. Earth Syst. Sci., 13, 17–25, 2009.

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, Hydrol. Earth Syst. Sci., 21, 4681–4691, doi:10.5194/hess-21-4681-2017, 2017.

Beniston, M. and Jungo, P.: Shifts in the distributions of pressure, temperature and moisture and changes in the typical weather patterns in the Alpine region in response to the behavior of the North Atlantic Oscillation, Theor. Appl. Climatol., 71, 29–42, doi:10.1007/s704-002-8206-7, 2002.

Bock, A. R., Hay, L. E., McCabe, G. J., Markstrom, S. L., and Atkinson, R. D.: Parameter regionalization of a monthly water balance model for the conterminous United States, Hydrol. Earth Syst. Sci., 20, 2861–2876, doi:10.5194/hess-20-2861-2016, 2016.

Bruno Soares, M. and Dessai, S.: Exploring the use of seasonal climate forecasts in Europe through expert elicitation, Clim. Risk Manage., 10, 8–16, doi:10.1016/j.crm.2015.07.001, 2015.

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., Déqué, M., Domeisen, D. I. V., Fröhlich, K., Hendon, H., Imada, Y., Ishii, M., Iza, M., Karpechko, A. Y., Kumar, A., MacLachlan, C., Merryfield, W. J., Müller, W. A., O'Neill, A., Scaife, A. A., Scinocca, J., Sigmond, M., Stockdale, T. N., and Yasuda, T.: The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter?, Quart. J. Roy. Meteor. Soc., 142, 1413–1427, doi:10.1002/qj.2743, 2016.

Chahine, M. T.: GEWEX: The global energy and water cycle experiment, Eos, Transactions American Geophysical Union, 73, 9–14, 1992.

Chimani, B., Matulla, C., Böhm, R., and Hofstätter, M.: A new high resolution absolute temperature grid for the Greater Alpine Region back to 1780, Int. J. Climatol., 33, 2129–2141, doi:10.1002/joc.3574, 2013.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 20, 3601–3618, doi:10.5194/hess-20-3601-2016, 2016.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Resour. Plann. Manage., 111, 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. L.: Seasonal climate predictability and forecasting: Status and prospects, WIREs Clim. Change, 4, 245–268, doi:10.1002/wcc.217, 2013.

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W. A., and Baehr, J.: Seasonal Predictability over Europe Arising from El Niño and Stratospheric Variability in the MPI-ESM Seasonal Prediction System, J. Climate, 28, 256–271, doi:10.1175/JCLI-D-14-00207.1, 2015.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, Geophys. Res. Lett., 41, 5620–5628, doi:10.1002/2014gl061146, 2014.

Finger, D., Heinrich, G., Gobiet, A., and Bauder, A.: Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss Alps and the subsequent effects on hydropower production during the 21st century, Water Resour. Res., 48, doi:10.1029/2011WR010733, 2012.

Förster, K., Meon, G., Marke, T., and Strasser, U.: Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany), Hydrol. Earth Syst. Sci., 18, 4703–4720, doi:10.5194/hess-18-4703-2014, 2014.

Förster, K., Oesterle, F., Hanzer, F., Schöber, J., Huttenlau, M., and Strasser, U.: A snow and ice melt seasonal prediction modelling system for Alpine reservoirs, Proc. Int. Assoc. Hydrol. Sci., 374, 143–150, doi:10.5194/piahs-374-143-2016, 2016.

Fundel, F., Jörg-Hess, S., and Zappa, M.: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices, Hydrol. Earth Syst. Sci., 17, 395–407, doi:10.5194/hess-17-395-2013, 2013.

Greatbatch, R. J., Gollan, G., Jung, T., and Kunz, T.: Factors influencing Northern Hemisphere winter mean atmospheric circulation anomalies during the period 1960/61 to 2001/02, Quart. J. Roy. Meteor. Soc., 138, 1970–1982, doi:10.1002/qj.1947, 2012.

Greenslade, D. and Berkhout, F.: Future Earth-Research for Global Sustainability, in: EGU General Assembly Conference Abstracts, vol. 16, 2014.

Hanzer, F., Marke, T., and Strasser, U.: Distributed, explicit modeling of technical snow production for a ski area in the Schladming region (Austrian Alps), Cold Reg. Sci. Technol., 108, 113–124, doi:10.1016/j.coldregions.2014.08.003, 2014.

Hanzer, F., Helfricht, K., Marke, T., and Strasser, U.: Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments, The Cryosphere, 10, 1859–1881, doi:10.5194/tc-10-1859-2016, 2016.

Hanzer, F., Förster, K., Nemec, J., and Strasser, U.: Projected cryospheric and hydrological impacts of 21st century climate change in the Ötztal Alps (Austria) simulated using a physically based approach, Hydrol. Earth Syst. Sci. Discuss., in review, 1–34, doi:10.5194/hess-2017-309, 2017.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Ineson, S. and Scaife, A. A.: The role of the stratosphere in the European climate response to El Niño, Nat. Geosci., 2, 32–36, doi:10.1038/ngeo381, 2008.

Jörg-Hess, S., Griessinger, N., and Zappa, M.: Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas, J. Hydrometeorol., 16, 2169–2186, 2015.

Kang, D., Lee, M.-I., Im, J., Kim, D., Kim, H.-M., Kang, H.-S., Schubert, S. D., Arribas, A., and MacLachlan, C.: Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems, Geophys. Res. Lett., 41, 3577–3585, doi:10.1002/2014GL060011, 2014.

Kaser, G., Großhauser, M., and Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes, Proceedings of the National Academy of Sciences, 107, 20 223–20 227, doi:10.1073/pnas.1008162107, 2010.

Kim, H.-M., Webster, P. J., and Curry, J. A.: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, Climate Dyn., 39, 2957–2973, doi:10.1007/s00382-012-1364-6, 2012.

Kirtman, B. and Pirani, A.: The State of the Art of Seasonal Prediction: Outcomes and Recommendations from the First World Climate Research Program Workshop on Seasonal Prediction, Bulletin of the American Meteorological Society, 90, 455–458, doi:10.1175/2008BAMS2707.1, 2009.

Kirtman, B., Power, S. B., Adedoyin, J. A., Boer, G. J., Camilloni, I., Doblas-Reyes, F. J., Fiore, A. M., Kimoto, M., Meehl, G. A., Prather, M., Sarr, A., Schar, C., Sutton, R., van Oldenborgh, G. J., Vecchi, G., and Wang, H. J.: Near-term climate change: projections and predictability, in: Climate change 2013 : the physical science basis : contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, 2014.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31, 13–24, doi:10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424-425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Krajči, P., Kirnbauer, R., Parajka, J., Schöber, J., and Blöschl, G.: The Kühtai data set: 25 years of lysimetric, snow pillow, and meteorological measurements, Water Resour. Res., 53, 5158–5165, doi:10.1002/2017WR020445, 2017.

Kumar, A., Chen, M., and Wang, W.: Understanding Prediction Skill of Seasonal Mean Precipitation over the Tropics, J. Climate, 26, 5674–5681, doi:10.1175/JCLI-D-12-00731.1, 2013.

Mackay, J., Jackson, C., Brookshaw, A., Scaife, A., Cook, J., and Ward, R.: Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom, J. Hydrol., 530, 815 – 828, doi:10.1016/j.jhydrol.2015.10.018, 2015.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): A High-Resolution Seasonal Forecast System, Quart. J. Roy. Meteor. Soc., 141, 1072–1084, doi:10.1002/qj.2396, 2015.

Marke, T., Strasser, U., Hanzer, F., Stötter, J., Wilcke, R. A. I., and Gobiet, A.: Scenarios of Future Snow Conditions in Styria (Austrian Alps), J. Hydrometeorol., 16, 261–277, doi:10.1175/JHM-D-14-0035.1, 2015.

Marzeion, B. and Nesje, A.: Spatial patterns of North Atlantic Oscillation influence on mass balance variability of European glaciers, Cryosphere, 6, 661–673, doi:10.5194/tc-6-661-2012, 2012.

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, Cryosphere, 6, 1295–1322, doi:10.5194/tc-6-1295-2012, 2012.

McCabe, G. J. and Markstrom, S. L.: A Monthly Water-Balance Model Driven By a Graphical User Interface, U.S. Geological Survey Open-File report, https://pubs.usgs.gov/of/2007/1088/, accessed on 22 Feb 2017, 2007.

Molteni, F., Stockdale, T. N., and Vitart, F.: Understanding and modelling extra-tropical teleconnections with the Indo-Pacific region during the northern winter, Climate Dyn., 45, 3119–3140, doi:10.1007/s00382-015-2528-y, 2015.

Montanari, A., Young, G., Savenije, H., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrolog. Sci. J., 58, 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

Olefs, M., Fischer, A., and Lang, J.: Boundary Conditions for Artificial Snow Production in the Austrian Alps, J. Appl. Meteorol. Climatol., 49, 1096–1113, doi:10.1175/2010JAMC2251.1, 2010.

Pagano, T., Garen, D., and Sorooshian, S.: Evaluation of official western US seasonal water supply outlooks, 1922-2002, J. Hydrometeorol., 5, 896–909, doi:10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2, 2004.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., and Winstral, A.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sens. Environ., 184, 139–152, doi:10.1016/j.rse.2016.06.018, 2016.

Pomeroy, J., Bernhardt, M., and Marks, D.: Water resources: Research network to track alpine water, Nature, 521, 32–32, doi:10.1038/521032c, correspondence, 2015.

Reid, W. V., Chen, D., Goldfarb, L., Hackmann, H., Lee, Y. T., Mokhele, K., Ostrom, E., Raivio, K., Rockstrom, J., Schellnhuber, H. J., and Whyte, A.: Earth System Science for Global Sustainability: Grand Challenges, Science, 330, 916–917, doi:10.1126/science.1196263, 2010.

Riddle, E. E., Butler, A. H., Furtado, J. C., Cohen, J. L., and Kumar, A.: CFSv2 ensemble prediction of the wintertime Arctic Oscillation, Climate Dyn., 41, 1099–1116, doi:10.1007/s00382-013-1850-5, 2013.

Robertson, D. E. and Wang, Q. J.: A Bayesian Approach to Predictor Selection for Seasonal Streamflow Forecasting, J. Hydrometeorol., 13, 155–171, doi:10.1175/JHM-D-10-05009.1, 2012.

Robertson, D. E., Pokhrel, P., and Wang, Q. J.: Improving statistical forecasts of seasonal streamflows using hydrological model output, Hydrol. Earth Syst. Sci., 17, 579–593, doi:10.5194/hess-17-579-2013, 2013.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, J. Climate, 27, 2185–2208, doi:10.1175/JCLI-D-12-00823.1, 2014.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519, doi:10.1002/2014GL059637, 2014.

Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H., Eade, R., Gordon, M., MacLachlan, C., Martin, N., Dunstone, N., and Smith, D.: Seasonal winter forecasts and the stratosphere, Atmos. Sci. Lett., 17, 51–56, doi:10.1002/asl.598, 2016.

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., Martin, N., Peterson, K. A., Rowlands, D., Carroll, E. B., Belcher, S., and Slingo, J.: Tropical rainfall, Rossby waves and regional winter climate predictions, Quart. J. Roy. Meteor. Soc., 143, 1–11, doi:10.1002/qj.2910, 2017.

Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrol. Processes, 21, 2075–2080, doi:10.1002/hyp.6825, 2007.

Schaefli, B., Hingray, B., and Musy, A.: Climate change and hydropower production in the Swiss Alps: quantification of potential impacts and related modelling uncertainties, Hydrol. Earth Syst. Sci. Discuss., 11, 1191–1205, 2007.

Schattan, P., Baroni, G., Oswald, S. E., Schöber, J., Fey, C., Kormann, C., Huttenlau, M., and Achleitner, S.: Continuous monitoring of snowpack dynamics in alpine terrain by aboveground neutron sensing, Water Resour. Res., 53, 3615–3634, doi:10.1002/2016WR020234, 2017.

Scherrer, S. C., Appenzeller, C., and Laternser, M.: Trends in Swiss Alpine snow days: The role of local- and large-scale climate variability, Geophys. Res. Lett., 31, L13 215, 1–4, doi:10.1029/2004GL020255, 2004.

Schick, S., Rössler, O., and Weingartner, R.: Saisonale Abflussprognosen für mittelgroße Einzugsgebiete in der Schweiz: Möglichkeiten und Grenzen hydrologischer Persistenz (Seasonal runoff predictions for mesoscale catchments in Switzerland – Potentials and limitations of hydrologic persistence), Hydrol. Wasserbewirtsch., 59, 59–67, doi:10.5675/HyWa_2015,2_2, 2015.

Schöber, J., Achleitner, S., Kirnbauer, R., Schöberl, F., and Schönlaub, H.: Impact of snow state variation for design flood simulations in glacierized catchments, Adv. Geosci., 31, 39–48, doi:10.5194/adgeo-31-39-2012, 2012.

Schöber, J., Schneider, K., Helfricht, K., Schattan, P., Achleitner, S., Schöberl, F., and Kirnbauer, R.: Snow cover characteristics in a glacierized catchment in the Tyrolean Alps - Improved spatially distributed modelling by usage of Lidar data, J. Hydrol., doi:10.1016/j.jhydrol.2013.12.054, 2014.

Schöber, J., Achleitner, S., Bellinger, J., Kirnbauer, R., and Schöberl, F.: Analysis and modelling of snow bulk density in the Tyrolean Alps, Hydrol. Res., 47, 419–441, doi:10.2166/nh.2015.132, 2016.

Sene, K., Piper, B., Wykeham, D., McSweeney, R. T., Tych, W., and Beven, K.: Long-term variations in the net inflow record for Lake Malawi, Hydrol. Res., doi:10.2166/nh.2016.143, 2016.

Sivapalan, M., Savenije, H. H. G., and Blöschl, G.: Socio-hydrology: A new science of people and water, Hydrol. Processes, 26, 1270–1276, doi:10.1002/hyp.8426, 2012.

Smith, D. M., Scaife, A. A., and Kirtman, B. P.: What is the current state of scientific knowledge with regard to seasonal and decadal forecasting?, Environ. Res. Lett., 7, 015 602, doi:10.1088/1748-9326/7/1/015602, 2012.

Svensson, C., Brookshaw, A., Scaife, A. A., Bell, V. A., Mackay, J. D., Jackson, C. R., Hannaford, J., Davies, H. N., Arribas, A., and Stanley, S.: Long-range forecasts of UK winter hydrology, Environ. Res. Lett., 10, 064 006, doi:10.1088/1748-9326/10/6/064006, 2015.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geogr. Rev., 38, 55–94, doi:10.2307/210739, 1948.

Trinh, B. N., Thielen-del Pozo, J., and Thirel, G.: The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems, Atmospheric Science Letters, 14, 61–65, doi:10.1002/asl2.417, 2013.

Twedt, T. M., Schaake Jr, J. C., and Peck, E. L.: National Weather Service extended streamflow prediction, in: Proceedings Western Snow Conference, pp. 52–57, 1977.

Vaughan, C. and Dessai, S.: Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework, Wiley Interdiscip. Rev. Clim. Change, 5, 587–603, doi:10.1002/wcc.290, 2014.

Viviroli, D., Weingartner, R., and Messerli, B.: Assessing the hydrological significance of the world's mountains, Mt. Res. Dev., 23, 32–40, 2003.

Warner, T. T.: Numerical weather and climate prediction, Cambridge University Press, Cambridge; New York, 2011.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, Journal of The Royal Society Interface, 11, 20131 162–20131 162, doi:10.1098/rsif.2013.1162, 2014.

Wilks, D. S.: Statistical methods in the atmospheric sciences, no. v. 91 in International geophysics series, Academic Press, Amsterdam; Boston, 2nd ed edn., 2006.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, L14 401, 5–5, doi:10.1029/2008GL034648, 2008.

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., and Clark, M.: Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill, J. Hydrometeorol., 17, 651–668, doi:10.1175/JHM-D-14-0213.1, 2016.

Yuan, X., Wood, E. F., Roundy, J. K., and Pan, M.: CFSv2-Based Seasonal Hydroclimatic Forecasts over the Conterminous United States, J. Climate, 26, 4828–4847, doi:10.1175/JCLI-D-12-00683.1, 2013.

Yuan, X., Wood, E. F., and Ma, Z.: A review on climate-model-based seasonal hydrologic forecasting: Physical understanding and system development, WIREs Water, 2, 523–536, doi:10.1002/wat2.1088, 2015.

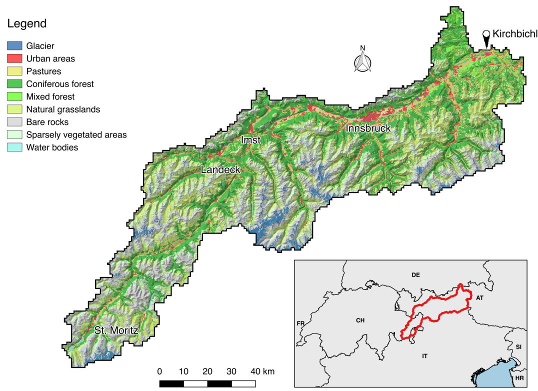

Figure 1. Map of the Inn headwaters upstream of the Kirchbichl gauging station. Major land use classes are shown along with the topography.

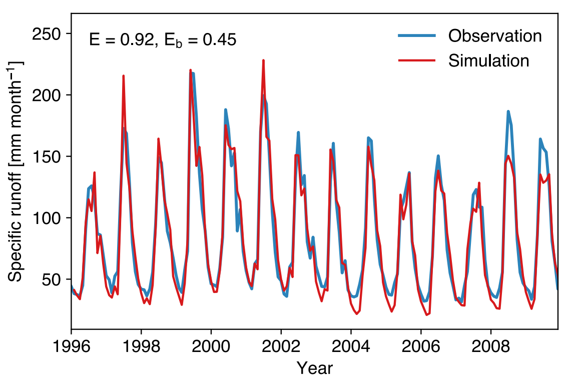

Figure 2. Specific runoff hydrographs for observed and modelled runoff at Kirchbichl gauging station and the corresponding performance measures E (Nash-Sutcliffe model efficiency, NSE) and E_b (benchmark NSE).

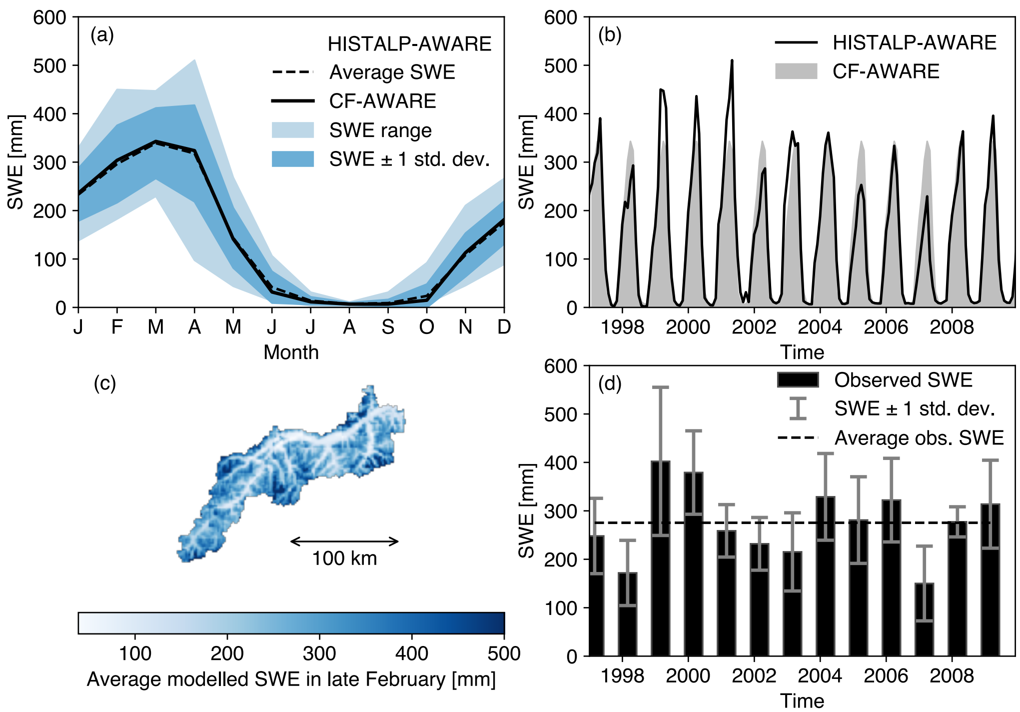

Figure 3. Observed and simulated snow states. (a) Range and standard deviation of SWE time series computed for single years and the corresponding mean monthly averages derived from AWARE using HISTALP data (HISTALP-AWARE). Average SWE conditions computed using climatological historical data forecasts (CF-AWARE) are displayed as well. (b) This plot shows the same data but as time series between 1997-2009. (c) Map showing the average spatial SWE distribution computed for February. (d) SWE observations in February collected from stations in the study area above 1400 m a.s.l. (Schöber et al., 2016). The error bars show the ± 1 standard deviation of observations for each year (see Schöber et al., 2016, for further details on SWE sampling).

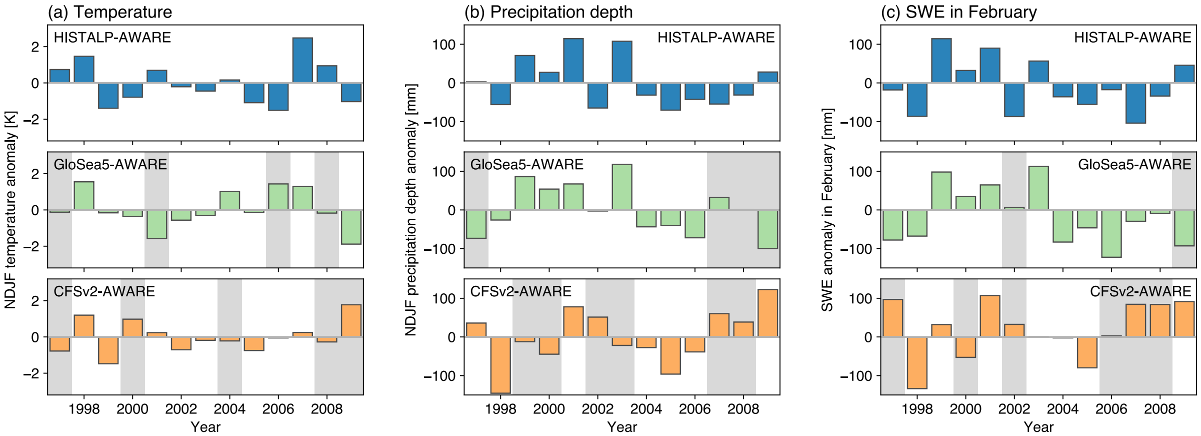

Figure 4. Anomalies of (a) basin-scale NDJF temperature, (b) NDJF areal precipitation depth, and (c) snow accumulation in February. Water balance simulations driven by CM-based seasonal forecasts are compared to water balance simulation driven by HISTALP. Shaded areas indicate years in which the sign of anomalies does not match the reference run.

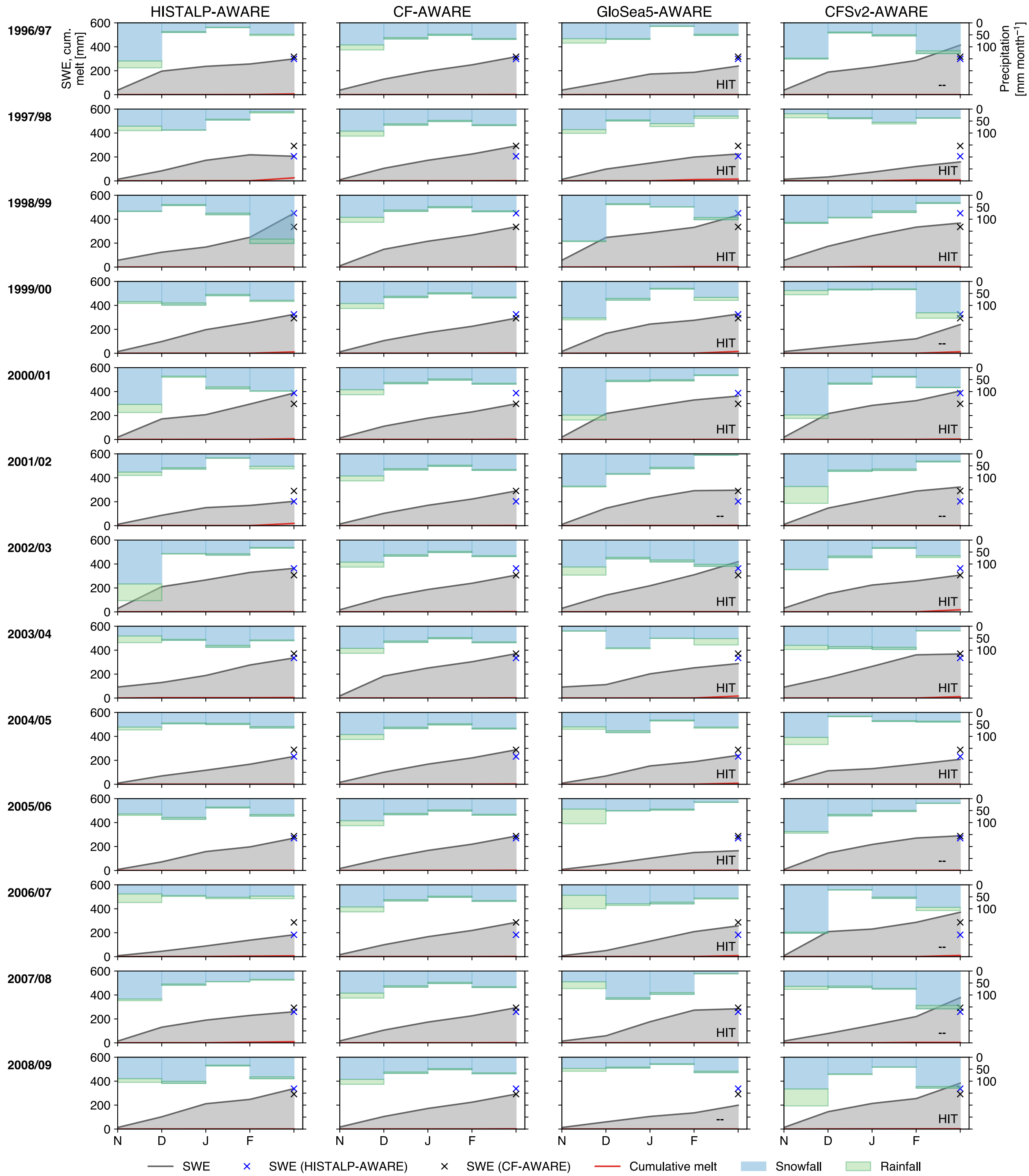

Figure 5. Water balance of the basin-scale snow storage for each year and each forcing dataset used for AWARE simulations. CF is the climatological forecast (longterm averages of HISTALP) which can be viewed as forecast yielding average conditions. The evolution of snow accumulation is categorized either "HIT" or "–" if the sign of anomalies obtained from HISTALP-AWARE and CM-based AWARE runs matches or mismatches, respectively.

Table 1. Skill measures for basin-scale averages of NDJF temperature, NDJF precipitation, and SWE in February. For each experiment, results for AWARE-GloSea5 and AWARE-CFSv2 are summarised. The first experiment (full dynamical model runs) refers to the standard setting of CM-based seasonal forecasts using AWARE (Sect. 3.2), while the other two experiments also involve a replacementof CM-based time series by climatology (Sect. 3.3). Skill measures: r: Pearson correlation coefficient, Std.: standard deviation of time series, RMSE: Root Mean Square Error, Hitrate: number of correctly predicted states (out of 13), BSS: Brier Skill Score, MAESS: Mean Absolute Error Skill Score. Please note: The units of Std. and RMSE are those of the input time series.

	Full dynamical model runs		Temperature from climatology		Precipitation from climatology	
	AWARE-GloSea5	AWARE-CFSv2	AWARE-GloSea5	AWARE-CFSv2	AWARE-GloSea5	AWARE-CFSv2
Temperature						
r [-]	0.32	0.17	-	-	0.32	0.17
Std. [K]	1.03	0.86	0.00	0.00	1.03	0.86
RMSE [K]	1.29	1.32	1.16	1.16	1.29	1.32
Hitrate [-]	9	8	-	-	9	8
BSS [-]	0.69	0.62	0.00	0.00	0.69	0.62
MAESS [-]	-0.02	-0.12	0.00	0.00	-0.02	-0.12
Precipitation						
r [-]	0.61	0.31	0.61	0.31	-	-
Std. [mm]	64.77	70.95	64.77	70.95	0.00	0.00
RMSE [mm]	55.97	78.66	55.97	78.66	62.20	62.20
Hitrate [-]	9	7	9	7	-	-
BSS [-]	0.69	0.54	0.69	0.54	0.00	0.00
MAESS [-]	0.19	-0.29	0.19	-0.29	0.00	0.00
SWE						
r [-]	0.57	0.28	0.44	0.25	0.51	0.16
Std. [mm]	72.57	71.91	62.69	68.03	23.01	20.32
RMSE [mm]	65.27	88.02	68.88	90.03	59.32	66.99
Hitrate [-]	11	7	10	7	8	6
BSS [-]	0.85	0.54	0.77	0.54	0.62	0.46
MAESS [-]	0.14	-0.23	0.06	-0.19	0.13	-0.05