We thank the anonymous reviewer #3 for doing excellent work and providing very useful comments, which will surely help us to improve the overall quality of the manuscript. In the following, we will scrupulously address and reply to each of his/her comments. We will include the necessary changes in the revised version of the manuscript.
Excerpts of other publications are quoted in quotation marks. Sometimes we refer to the "first paper", which is our earlier publication on this subject (Kormann et al., 2014). The referee comment is written in blue colour, our answers to the comments in black.

The paper of Kormann et al. addresses a relevant scientific question within the scope of HESS. It analyzes streamflow trends in an alpine region and attempts to explain to which extent the observed changes are caused by changes in climate variables. The intelligent combination of different methods and process understanding allows the authors to formulate and support hypotheses. For example, the argument that annual trend analyses may not be informative due to the integration of counteracting processes within the annual period is convincing and supported by their results. I also appreciate very much the efforts to introduce process understanding in the design of the study and in the analysis of the results. Very illustrative are also Figures 8 and 9, schematically summarizing streamflow changes and the associated drivers. The paper is a substantial contribution to answering the question of hydrological change in mountain areas. Its novelty lies particularly in the smart combination of different methods. It addresses detection and attribution of change at the same time and advance further than many other papers on hydrological change. Overall, I am very positive and recommend publication in hess.

General comments:

Besides a number of specific and technical comments, I have the following major criti-sisms. The presentation of the methods and result is rather "dense", and there is overlap with another paper (Kormann et al., in press):

(1) Explanations do not suffice to understand the methods and one could not redo this analysis without reading a number of other papers. I understand that the paper would get very long if all the methods would be given in detail, but I feel that more information on the methods should be given. I have made a few proposals where I feel that additional information would be very good.

Thanks for this comment. We will consider each one of the referees proposals in the revised manuscript. The method section will be generally revised to facilitate understanding, while still trying to keep the manuscript short.

(2) To understand the results, the reader must pay close attention not to get lost. The paper is not an easy read. I wonder if the authors could facilitate reading this paper by adding more explanations and guiding the reader more smoothly through the material.

Thanks for the comment. We will further improve the manuscript as the referee recommends (e.g. with a schematic illustration on the different methods used, see Appendix, or a restructuring of the Methods section, as Referee #4 proposed). We will try our best to add more explanations and further guide the reader through the manuscript.

(3) There seems to be quite some overlap with another paper (Kormann et al., in press) from the first author, dealing more or less with the same data/region. In several instances the reader is referred to the other paper (which is not yet available), so understanding is sometimes difficult. Further, the question arises how novel the hessd paper is. I cannot answer this question since I do not know the other Kormann paper. The hessd paper should be written in a way that it is understandable on its own and that its contribution is very clear.

We answered to this point already in a separate comment (http://www.hydrol-earth-syst-sci-discuss .net/11/C2850/2014/hessd-11-C2850-2014- supplement.pdf).

Specific comments:

p6883-24: Are these metrics (centre of volume, day of occurrence of the annual peak flow) more sensitive than, for example, streamflow volume, quantiles etc.? If yes, please provide an explanation.

These metrics are used as a simple proxy for indicating effects of climate change on alpine streamflow (e.g. earlier snowmelt). We will clarify this.

p6886-Data: The temperature and snow height stations used in the paper are never shown. I propose to add these stations to Fig. 1 or add another figure showing them.

Thanks, we will include a map with the T and SH stations in the revised manuscript.

p6886-17: "... The number of stations is a trade-off between a large number of stations that cannot be interpreted in a detailed way and an insufficient number of stations that cannot be rated as representative...". This sentence may be true, but what is the purpose of this statement? Does this mean that you have selected only a part of the available streamflow (temperature, snow height) stations? If yes, please give more information on which basis you have done the selection. How have you determined which sub-set of stations is representative?

Thanks for the comment, we will clarify this. We intended to perform a regional trend analysis, for this purpose we selected the stations. This is contrary to the majority of studies, which analyse trends for a larger area (e.g. nation-wide, european-wide, US-wide). But with more stations, it is more challenging to interpret the results thoroughly, as different hydroclimatological conditions could potentially mask and thus complicate finding clear and coherent trend patterns. Moreover, most trend studies only describe and interpret the spatial variability of the 3-monthly or annual trends, which is a too coarse solution in our opinion. We think, these approaches are responsible for the fact that many trend studies, even in mountain regions (where climate change signals should be stronger), often reveal “inconclusive or misleading findings” (Viviroli et al., 2011).
The aim of the overall project was to look at streamflow changes in North Tyrol, which primarily determined our selection. However, as we finally had to exclude many discharge gauges as they

were influenced by hydro power, there were (in our opinion) not enough datasets to provide representative statements. For this reason, we included further gauges in the surrounding area within an additional range of approx. 40 km.

p6886-23: The decision not to study precipitation trends needs a clearer explanation. There seem to be 3 justifications: (1) "... precipitation did not reveal any clear trend patterns ...", (2) "... snow height changes have a much stronger effect on streamflow than those of snowfall ...", (3) "... we assume that precipitation has no trend. The validity of this assumption is supported by the fact that precipitation changes are most probably of a far smaller magnitude than changes caused by e.g. increased glacial melt ...". I find this difficult to understand. What exactly made you decide to refrain from analysing precip trends? Why do you assume that precip has no trend when precip did not reveal any clear trend patterns? Do you speak about regional precip trends / spatially coherent precip trends?

We analysed precipitation trends in the earlier paper (Kormann et al., 2014) and we could not find any significant trend patterns in precipitation, which was also reported in other studies. Some significant trends were found but these were spatially not coherent.
This means, spatially incoherent trends possibly might exist, but they cannot be detected due to a low signal-to-noise ratio. Anyway, if these trends would exist, there would probably not be a clear signal in streamflow trends as there is no homogeneous signal in the precipitation trends. Lastly, the trends in precipitation that we found were small in magnitude compared to the streamflow trends. We will add an explanation to clarify this issue.

The sentence "... precipitation changes are most probably of a far smaller magnitude than changes caused by e.g. increased glacial melt ..." is not clear. Do you mean 'changes in streamflow caused by increased glacial melt'?

Yes, we will correct this.

p6888-14: Please give more explanations about the prewhitening methods you apply "... prewhitening methods described in Wang and Swail (2001) were applied ...". Did you apply several methods? Or just prewhitening for lag 1?

We just prewhitened for lag-1-autocorrelation, as usual. We will add further explanations.

p6889-Equation 1: I do not understand equation 1 and feel that the explanation of MDT is not comprehensive enough. It would be good if one could understand MDT without going to Morin (2011). How generic is this equation? Does it apply to linear trends only? Has Morin (2011) used certain distributions in his Monte Carlo experiment and would this limit the application of MDT? Further, I am not sure what MDT adds to the work. From Fig. 2 I learn that trends are significant when they are outside the MDT band. If this is the case, then what additional information does MDT give?

The MDT points out the role of the signal-to-noise ratio when detecting trends. It makes visible, what otherwise might not have been obvious: That only at stations, where the detected trends are higher than a certain level (which is determined by the variability and the record length), the trends are significant. The MDT helps to support our 3rd hypothesis: Trends in mid-altitudes are not detected due to (1) the high variability in the data and (2) the low signal, which is caused by a compensating effect of increased glacial melt in higher altitudes and increasing ETP at lower altitudes. We will further explain this and additionally improve the comprehensibility of the equation.

p6889-section 3.1.3: Again, I think that more information about the method should be presented.

Thanks for the comment, we will add further information.

p6892-9: Do you only average Tmin over all stations? If yes, does this mean that Tmin behaves similar across all stations but not Tmean and Tmax? What is the explanation for this result?

We needed some adaption of T trends to the mean watershed heights. As we found out that T trends in general (also Tmean and Tmax (!)) behave similarly across most of the stations analysed (Fig. 5 a)-c)), we averaged the daily trends over all stations analysed. However, Tmin proved to be most beneficial for the multiple regression model, so we only mentioned Tmin. We will clarify this in the revised manuscript.

p6895-11: I do not understand the following sentences: "... The Mann–Kendall trend test has been criticised in some recent publications, particularly for the following issues: streamflow is usually not an independent and identically distributed variable, which is a precondition for using the MK test. Furthermore, a trend could be nonlinear or a part of a multispectral oscillation. Therefore, similar to Déry et al. (2009), the Sen's Slope Estimators are presented as well without assigning trend significance. ..." The Mann-Kendal test estimates the significance of gradual trends and Sen's slope estimates the magnitude/slope of a gradual trend. Hence, both methods give complementary information and are usually applied together. This is done also in this paper which is fine. However, the given justification is strange: (1) independence: this should have been considered via prewhitening, (2) nonlinear: the Mann-Kendall test does not require that the trend is linear, but it tests gradual change, (3) part of multispectral oscillation: I do not see that Sen's slope deals in a better way with oscillations.

We understand that a clarification is needed and that this section may be misleading. We actually wanted to question the use of the Mann-Kendall test (or significance tests) as such (which is also done in other trend studies, e.g. Déry et al., 2009). With this, we justify our decision of not using the Mann-Kendall test in the further analyses of the paper.
We did not aim to say that the Sen's Slope does the same thing or has better qualities.

p6914 - Caption Fig. 1: I feel that this figure needs more explanation (in particular, since the other Kormann paper is in press only). Please give the significance level used. What exactly means 'trend in percent'? Even stations with 1% trend are significant - this is somewhat surprising. What is the time period studied?

Technical corrections: Several locations: The reference "Kormann et al., 2013" needs to be corrected to "Kormann et al., 2014".
Several locations: Trend magnitudes are given in %. How are they calculated? Change in magnitude during 1980-2010 divided by mean magnitude?

Thanks for the comment, we will correct and clarify the corresponding sections. Trends are given in per cent change per year, with a significance level of *alpha*=0.1. The magnitudes in per cent are calculated from the *change per year* divided by *mean annual* streamflow. This is maybe why the magnitudes seem pretty small (but also trends of small magnitude may become significant when total variability is low). For knowing the change during the whole period, one has to multiply it with the number of years studied (31 years).

p6887-9: Does this sentence "... glacier mass balances have been completely negative only since the 1980s ..." refer to the Greater Alpine area?

Abermann et al. (2009) refer only to the Ötztal Alps. However, we will change it for Abermann et al. (2011) who reported about mostly negative mass balances since 1980 for whole Austria: "Since 1980, only a few years (e.g. 1984 or 1989) interrupt the generally negative trend." (Abermann et al., 2011)

p6889-5: Is Sen's slope really the "... mean of the slope between all possible pairs of data points ..."? I thought it was the median.

Thanks, we will correct that.

p6889-20: What do you mean with "... averaged observations ..."?

With "standard deviation of the series of averaged observations" we mean the standard deviation of a dataset, that already has been aggregated to a certain time resolution for analysing trends (e.g. the standard deviation of *annual* averages of temperature). We will clarify.

p6890-11: The acronym 30DMA should not be used in the section title because it is introduced later.

Thanks, we will correct it.

p6890-15: What do you mean by "... temporal relationship ..."? A relationship which changes in time?

Thanks for the comment, we will clarify. Simply said, we meant that if something happens in one of the predictor variables on a certain day of year (e.g. T crosses the freezing point in spring; T trends turn up; snow height has reached its maximum in winter), and trends in streamflow turn up as well around this day of year, then this might indicate the causes for the streamflow trends.

p6892-3: These possible predictor variables are the indicators for temperature (mean, min, max) and snow height, right? In the current version, this sentence is somewhat cryptic.

Thanks, we actually wanted to say that we tried different variables such as catchment properties, seasonal cycles of different variables, trends of different variables, etc., to find out which predictors (independent variables) could cause Q trends. We will revise this sentence.

p6896-8: I do not understand what you mean with 'Comparing single stations with each other' in the sentence "... Comparing single stations with each other, it is shown that the fieldsignificant T trends appear in clusters that start and end during similar DOYs ..." Field significance looks at the complete collection of stations, it does not compare single stations.

Thanks for the comment. We intended to say: Comparing single stations with each other, it is shown that analogue T trends appear in clusters that start and end during similar DOYs. We will change it accordingly.

p6896-23: Why should it be obvious? How do I know that snow height has a low signal-to-noise ratio?

We agree with the referee, we will clarify this. Simply said, there is only field significance, if there are enough significant trends at a certain DOY in the station datasets. Field significance is found at the end of winter, meaning that only during this time of year, there are enough significant trends. Variability of snow height is relatively higher than the one of e.g. temperature. For this reason, it has a low signal-to-noise ratio as well, because otherwise, more significant trends would have been detected.

p6901-6: Could you please extend the following sentences? I am not sure what is meant here: "... Our regression approach does not presume to capture the complete set of predictors, but is just meant as an heuristic approximation, as the Durbin–Watson statistic indeed indicates. Therefore, the coefficients should be taken with caution, since standard uncertainty measures cannot be derived in that case. ..."

We found some autocorrelation in the residuals (Durban Watson statistic = 1.4312). This is violating the assumptions of independence of linear regression, which often happens when fitting models to time series with a seasonal cycle. The autocorrelation in the residuals precludes statements on confidence bands and significance tests: The standard errors of the regression coefficients are potentially too small, which pretends higher model precision. However, our model stands as an approximation only. We are aware that the model is not perfect, as it is impossible to find all specific causes that explain the streamflow trends in our study region. The model is able to simulate streamflow trends sufficiently well, providing further hints on the causes of Q trends.

p6916-Fig3: Upper panel: I propose to change the color for 'not significant' from dark blue to a color (e.g. white) which is not used for coding magnitude.

We agree with the referee and we will change the coding accordingly.

p6917-Fig4: It seems that Figure 4 is not mentioned and discussed in the text.

On p6897-1 to -11 we mentioned Fig. 4:
"Besides the trends, we derived the characteristic dates of T and SH: The average DOYs of daily T_{min}, T_{max} and T_{mean} surpassing the freezing point ($DOY_{0°Tmean/max/min\_Spring/Autumn}$) all depend on station height, in spring as well as in autumn (Fig. 4a and b). The same applies for the average DOY of the annual snow height maximum (DOY_{SHmax}, Fig. 4c). Lines were fitted to represent these relationships. Nearly all the relationships analysed were found to be approximately linear. Only Fig. 4c shows that there might exist some height-independent DOY_{SHmax} at mid-altitudes. However, as these irregularities were not very strong, a linear relationship was also applied. The corresponding equations were used to transfer the respective DOYs ($DOY_{0°Tmean\_Spring}$, $DOY_{0°Tmean\_Autumn}$, ...) to the mean altitudes of the watersheds considered in this study."
However, we will further discuss this Figure in the revised version in a separate discussion section.

p6919-Fig6: Please include the line of perfect fit.

Thanks for the comment, we will consider this.

References:

Abermann, J., Lambrecht, A., Fischer, A., and Kuhn, M.: Quantifying changes and trends in glacier area and volume in the Austrian Ötztal Alps (1969–1997–2006), The Cryosphere, 3, 205–215, doi:10.5194/tc-3-205-2009, 2009.

Abermann, J., M. Kuhn, and A. Fischer. A reconstruction of annual mass balances of Austria's glaciers from 1969 to 1998. Annals of Glaciology 52.59, 127-134, 2011.

Kormann, C., Francke, T., and Bronstert, A. (2014), Detection of regional climate change effects on alpine hydrology by daily resolution trend analysis in Tyrol, Austria, J. Water Clim. Change, in press, 2014.

Appendix

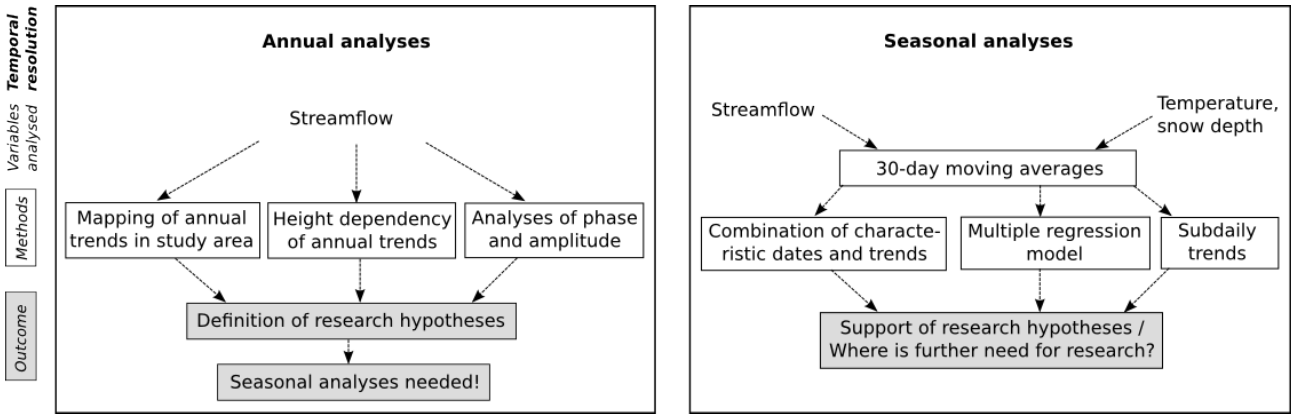

Figure 1: Schematic illustration on the approaches applied in the manuscript.