Reply to Anonymous Referee 1

We thank Referee 1 for reviewing our manuscript and providing constructive, actionable feedback. Below we provide our responses to each point raised.

Comment: P1, L8-16: The abstract should clearly include concrete results. The paper is far more interesting than what the abstract suggests. For example, a mention of the way the models are used in sequential mode, or the skill achieved, can be mentioned there.

Reply: This is an excellent comment, and we have added the following text to the abstract of the working manuscript (P1, L16-27):

"Skillful results (forecast outperforming climatology) are produced for short lead-times (September 1st; RPSS = 0.31), where categorical hit skill is 61%, with Above (wet) and Below-Normal (dry) flow years achieving 82% and 64% categorical hit skill, respectively. At longer lead-times, climatological skill exceeds forecast skill, largely due to less observations of precipitation in the statistical model (August 1st RPSS = 0.02 and July 1st RPSS = -0.39). Coupling the September 1st statistical forecast model with a Niño 3.4 region sea surface temperature phase and strength statistical model allows for equally skillful categorical streamflow forecasts to be produced for a May 1st lead, triggered for 60% of the years in the period 1950-2015. The reservoir allocation model is skillful at the September 1st lead (categorical hit skill score = 53%), and using a probabilistic modeling approach, forecast-based allocations are categorically skillful (79%) when the model predicts the observed allocation category with at least 80% confidence ($\geq$ 80% of annual forecast values fall within a single category). The frameworks applied here advance the understanding of the mechanisms and timing responsible for moisture transport to the Elqui Valley, and provide a unique application of streamflow forecasting in the prediction of per-water right allocations. Both have the potential to inform water right holder decisions."

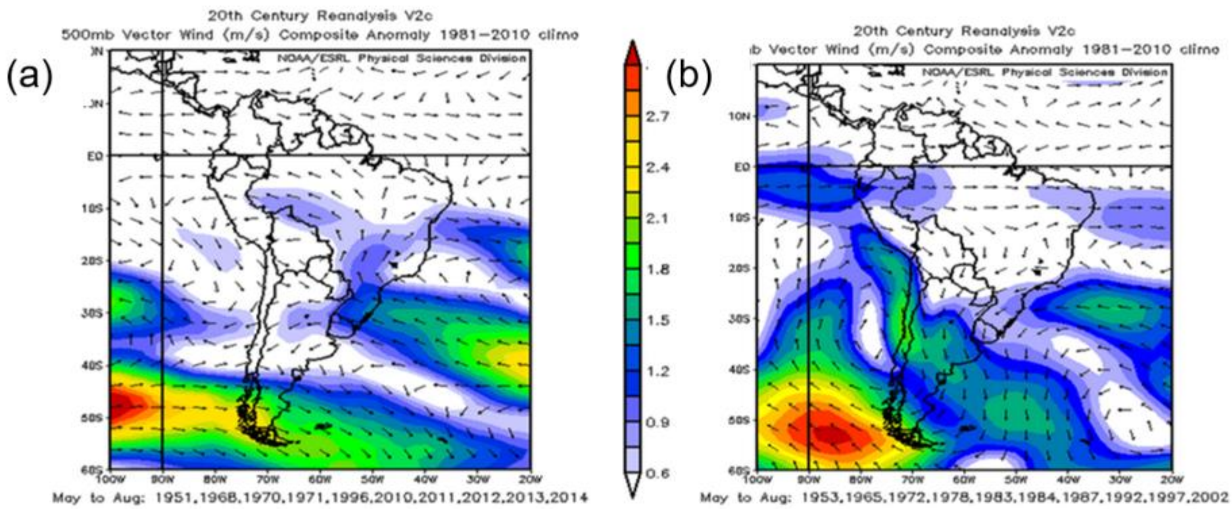
Comment: P7, Section 2.1: Is the data quality-controlled? Maybe add a sentence or two about that, so the reader knows if the data can be trusted.

Reply: Observations of streamflow and snow water equivalent are obtained from the Direccion General de Aguas (DGA), a department of the Ministry of Public Works of the Chilean Government. Collection, validation and quality control of hydrologic measurements are part of DGA's core functions; thus, we treat the data as fully vetted and having met DGA's quality control standards. The referee's comment is valid, and warrants an addition to the manuscript. We have added the following (P7 L15-18):

"One of DGA's primary functions as the regulator of surface water resources for the Chilean Government is to collect, validate, and perform quality control of hydrologic measurements. Open source data obtained through DGA is considered as having met DGA quality standards."

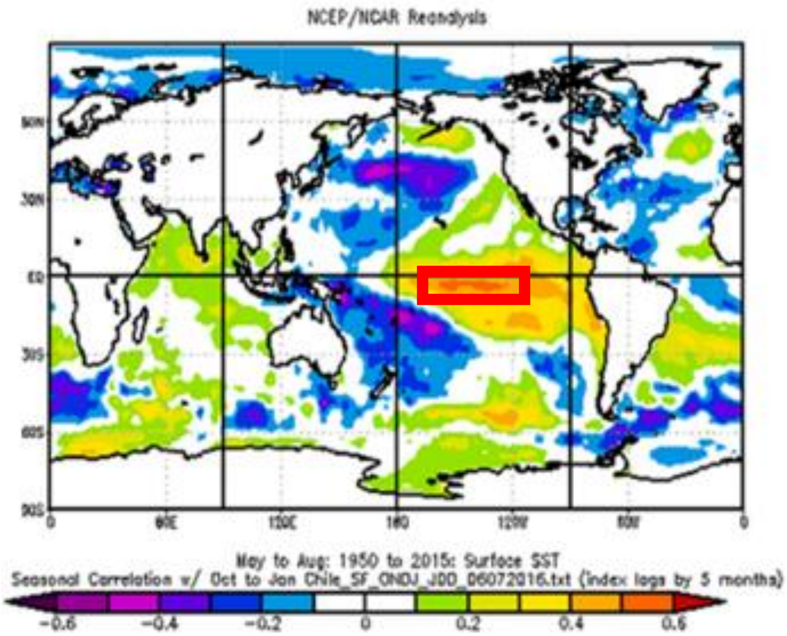
Comment: P7, L17, L21 (and other places): vector winds? This is the first time I see that name. What the authors mean by that? To use both u and v? Why do not just say “winds”?

Results: We are concerned with both the magnitude (colors) and direction (arrows are the resultant of u and v) of vector winds at 500mb (Fig. 4. (a) and (b), *excerpt below*). Both are critical in terms of determining the efficiency of moisture transport to the Elqui Valley. If we ignore the direction component, “winds” or “wind speed” should be used. Additionally, “vector winds” is the name commonly utilized by NOAA and other climate agencies.



Comment: P8L8: why to do a spatial average? I do not fully understand that sentence.

Reply: Gridded potential predictors are identified through spatial composite and/or correlation mapping (e.g. sea surface temperatures (SSTs). To extract the signal(s) from within the gridded data set and avoid noise present at the grid scale, principal component analysis (PCA) is commonly applied to the gridded data. Correlating the first principal component (PC), which is the strongest signal, with the spatially averaged data identifies whether the signal is spatially homogenous. Alternatively, if the first PC does not correlate well with the spatial average, the heterogeneity of the dataset is important, and thus using the spatial average may not be the best approach. For example, the spatial average of SSTs (Fig. 4 (c)), which is identified as a potentially significant as predictor of streamflow for the Elqui River (roughly consistent with the Niño 3.4 region), correlates highly (>0.9) with the first PC of the gridded SST data. Having identified SSTs as spatially homogenous, and consistent with the Niño 3.4 region, we correlate and ultimately select the Niño 3.4 Index as a potential predictor of streamflow as it is well-known, well understood, and well-studied. We do this as opposed to selecting a marginally different (e.g. sub-region of the Niño 3.4 region), but much less understood and perhaps less defensible area. Furthermore, using the spatial average rather than selecting a sub-regional area may be a more conservative approach as it does not guarantee that the strongest possible relationship is identified. In addition, using an index avoids grid cell selection bias (cherry-picking), which could result in an insufficient number of grid cells to be statistically significant, or produce vastly different regions of high correlation (spurious correlations.) The following papers support the claim of teleconnections between precipitation and SST in north-central Chile (Aceituno 1988; Falvey and Garreaud 2007; Garreaud et al. 2009; Montecinos and Aceituno 2003).



Comment: P9L1: are the authors talking about CFSR? They are talking about NCEP-NCAR reanalysis but then they cite Saha et al 2013. Is there a confusion here?

Reply: We appreciate the reviewer catching this improper citation. Clearly, the citation should be (Kalnay et al. 1996) as opposed to Saha et al 2013. We also add (Huang, van den Dool, and Georgarakos 1996) to specifically reference CPC's soil moisture data. The working manuscript has been updated with the appropriate citations.

Comment: P11L21: why a forecast is not issued in that case? Explain in the text.

Reply: The statistical phase and strength model (Stat-P&S) at the May lead does not provide a categorical streamflow forecast for Niño 3.4 Index = (+0.5°C, +0.75°C) or (-0.5°C, -0.75°C) as the range is considered transitional (not weak or moderate as identified by NOAA). Both the magnitude and persistence of SST observations in this range do not allow for production of skillful forecasts. The May 1st forecast lead uses January-April Niño 3.4 Index values to categorically forecast October-January streamflow, which requires prediction through the Spring Barrier (Duan and Wei 2013). Typically, SSTs within the transitional range are not stable and actively moving to either a neutral or strengthened phase. Until these changes occur, at some date beyond May 1st (typically beyond the Spring Barrier), a categorical or deterministic forecast is typically not skillful. Deferring to the September 1st statistical principal component regression model (Stat-PCR) is warranted when SSTs are in the transitional range.

The question is valid, and our original manuscript does not address the reasons for which the transitional range within the Niño 3.4 Index is not used by the Phase and Strength model at the May lead. To provide clarity in our approach we have added the following to the working manuscript (P11 L21-24):

"For these ranges, neither the magnitude (not weak or moderate as defined by NOAA) nor persistence of SST observations allow for production of skillful categorical streamflow forecasts. For years where SSTs fall within these ranges at forecast leads prior to the Spring Barrier, strength and phase are subject to rapid transition, and categorical forecasts are typically not skillful."

Comment: P11Section2.2: I suggest to change the title of the subsection, as it seems to be about a proper dynamical prediction model, and it is really about using dynamical model output in a statistical model.

Reply: Regarding dynamical model prediction, we initially considered raw dynamical climate model outputs of precipitation and SSTs to predict streamflow (since clearly streamflow is not an output of dynamical climate models), but the results were poor. We thus proceeded with statistical post-processing as a means of correcting dynamical model outputs (Gheti 2008). The reviewer's point is valid as the sub-section title may be interpreted as dynamically modeled streamflow, including a physically-based hydrology model. To avoid confusion and as a means of accurately describing the forecast approach we have changed the title of 2.2 to "Hybrid dynamical-statistical streamflow prediction model" to capture the fact that predictors come from the dynamical model, but the prediction model formulation is still statistical in nature.

Comment: P12L8: authors should be a bit more explicit about when local variables have predictive strength. Conditions? Dates? Proportion of total cases? More information is needed.

Reply: Local variables and their predictive strength are discussed in 2.1, and shown in Figure 5 (a) and (c) and in Table 1. The same variables are used, when appropriate, for the statistical model using corrected (quantile mapping) GCM outputs for precipitation and SSTs (Stat-Dyn), with forecasts issued January 1st , May 1st and June 1st. For these leads, local variables are not useful, and therefore only GCM predictions of precipitation and SSTs are used (Table 2. *Excerpt below*). This is not a surprising result considering local variables are skillful in the prediction of October-January streamflow only during months of peak precipitation (May-August) as shown in the manuscript in Figure 5 (a) and (c).

Still, we recognize readers may benefit from additional explanation and have added the following for clarity of local variable inclusion in the working manuscript (P12 L10-13):

"The Stat-Dyn model is meant to provide streamflow forecasts at extended leads, beyond what is possible with global and local observed data used to inform the Stat-PCR model. Local variables (e.g. precipitation, snow water equivalent and soil moisture) hold the most predictive strength during the season of peak precipitation (May-August) and thus are only considered for the Stat-Dyn model for leads at prior to June 1st (Fig. 5 (a.) and (c.)."

Forecast		Retained Predictors		
Statistical Approach (Stat-PCR)	Sep 1st	Aug SM	JA Prcp	Aug 3.4
	Aug 1st	Jul SM	JJ Prcp	Jul 3.4
	Jul 1st	Jun SM	MJ Prcp	Jun 3.4
Dynamical Approach (Stat-Dyn)	Jun 1st	JJA 1.2	JJA Prcp	-
	May 1st	JJA 3.4	JJA Prcp	-
	Jan 1st	JJA 3.4		-

Comment: P14L1-2: please check the syntax of the sentence.

Reply: We agree the structure of the sentence can be improved to better illustrate the point. We have changed the sentence in the working manuscript to:

"Allocation, as issued annually by JVRE, and storage outcomes are hindcast in a cross-validated mode for the period of record (1950 – 2015) by coupling the streamflow prediction models to a simple reservoir balance model."

Comment: P14Section2.4: I suggest to remind the reader that all these results are obtained using cross-validation (a lot of studies out there do not even bother to cross-validate!)

Reply: We thank the referee for the comment, and have included language which reminds the reader the forecast outputs are cross-validated.

Comment: P15L9: why Pearson coefficient?

Reply: Pearson's correlation coefficient is commonly used to assess both the general parametric association between forecast and observed values, and phase error. While it doesn't account for forecast bias and is sensitive to outliers, it is selected because it is well known and well understood. In addition, we utilize RPSS and categorical skill score metrics which describe additional performance and features of the forecasts.

Comment: P19L20: approach or model? Which one?

Reply: We appreciate the referee noticing and highlighting this error. For consistency, we use "model".

Comment: P19L20-21: I do not understand the sentence. When the other 40% occur?

Reply: 40% refers to a fraction of the number of years in the record (1950-2015) not predicted by the Stat-P&S model at the May lead using January-April Niño 3.4 Index because the index values fall within the transitional ranges (+0.5°C, +0.75°C) or (-0.5°C, -0.75°C). The transitional ranges do not provide skillful categorical forecasts for the May 1st lead. For this reason we do not forecast these years until the Stat-PCR model is skillful for the September lead. Our coupled statistical prediction model defers prediction for these years to September.

Comment: P20Step2a and Step2b: what is the real difference here?

Reply: The difference between Step 2a and 2b relates to whether the Stat-P&S model issues a May forecast. If January-April Niño 3.4 region SSTs meet the Stat-P&S criteria, a May 1st categorical forecast is issued (Step 2a). Otherwise, the Stat-PCR model is used to produce a September 1st forecast (Step 2b). The novelty of coupling the Stat-P&S and Stat-PCR models is the Stat-P&S model provides an initial, categorical indication (May 1st lead) of October-January streamflow for 60% of years between 1950-2015. From the perspective of a water rights holder, a skillful categorical forecast at a May 1st lead may provide relevant information to inform October-January decision making (e.g. cropping decisions by water right holding farmers). The initial forecast is reinforced by the Stat-PCR model, which provides a skillful deterministic forecast, but only for a September 1st lead. The key is that the agreement between the May 1st categorical forecast produced by the Stat-P&S model (when issued) and the September 1st deterministic Stat-PCR model is very high. In applied terms, a water right holder seeking to make forecast informed decisions has information which outperforms climatology at leads up to 4 months prior to the issuance of the actual allocation value.

Comment: P25L25: 92% is extremely high. Can you please confirm there is not a typo there?

Reply: We agree, 92% is a very high 'hit score'. However, this is a calculation of how often the forecast category aligns with the observed category, and is not a correlation. The hit score referenced here is for two categories, divided at 0.75 L s^{-1} as opposed to three categories as discussed at length in the paper. This case provides less overall information (probability of the allocation above or below the threshold) and thus we are not overly surprised that the score increases dramatically. In fact, we select this case specifically to compare with the three-category allocation model, to illustrate how categorical skill is a product of the bounds selected by the stakeholder.

Comment: P14, eqs 3 and 4: how are these equations obtained?

Reply: The purpose of the reservoir allocation model is to compare allocation and storage outcomes from the forecast and climatology informed reservoir operations against observations (which constitute a perfect forecast); this provides a means of evaluating the streamflow forecast in an applied context. Thus, equations 3 and 4 are simply a modified version of the reservoir balance. To include the annual, end-of-year storage target used by reservoir operators in the Elqui (100 million cubic meters), we adjust allocation for period $i+1$ by the storage deficit or surplus at the end-of-year i. For example, if the forecast informed allocation, $A_{i_{ONDJ}}$, results in an end-of-year storage $S_{i_{Feb}} \leq 100M\ m^3$, $A_{i+1_{ONDJ}}$ is penalized by the difference $100 - S_{i_{Feb}}$. In contrast, if $S_{i_{Feb}} \geq 100M\ m^3$, $A_{i+1_{ONDJ}}$ is boosted by the absolute value of the difference of $100 - S_{i_{Feb}}$. It is important to note that the equations are applied uniformly to the forecast, climatology, and observations, so a fair assessment of performance can be achieved.

Comment: P16L22-24: confusing sentence… too many commas?

Reply: We thank the referee for this comment, and agree that the sentence is confusing. We have replaced it the working manuscript with the following:

"As forecast lead increases, both Hit Score and RPSS decrease, while Extreme Miss Score increases. These results occur because less information regarding the MJJA rainy season is available, which is consistent with decreased correlations between ONDJ streamflow and predictors (Fig. 5.)"

Minor Comments: (all accepted and corrected in the manuscript)
P1L20: "institutions"

P4, L13, L16: something is wrong with the way the references are being written. E.g., it should be Aceituno, 1988.

P10: define N.

P15L10: …as opposed to *a specific quantity*….

P11L28: I think the authors mean "North American Multi-Model Ensemble".

P16L2: references in capital

P17L11: maybe change "affirming" to "confirming"?

References:

Aceituno, Patricio
1988 On the Functioning of the Southern Oscillation in the South American Sector. Part I: Surface Climate. Monthly Weather Review 116(3): 505–524.

Duan, Wansuo, and Chao Wei
2013 The "spring Predictability Barrier" for ENSO Predictions and Its Possible Mechanism: Results from a Fully Coupled Model. Internation Journal of Climatology 33: 1280–1292.

Falvey, Mark, and René Garreaud
2007 Wintertime Precipitation Episodes in Central Chile: Associated Meteorological Conditions and Orographic Influences. Journal of Hydrometeorology 8(2): 171–193.

Garreaud, René D., Mathias Vuille, Rosa Compagnucci, and José Marengo
2009 Present-Day South American Climate. Palaeogeography, Palaeoclimatology, Palaeoecology 281(3–4). Long-Term Multi-Proxy Climate Reconstructions and Dynamics in South America (LOTRED-SA): State of the Art and Perspectives: 180–195.

Gheti, Rares
2008 Statistical Post-Processing of Dynamical Surface Air Temperature Seasonal Predictions Using the Leading Ocean-Forced Spatial Patterns. http://digitool.library.mcgill.ca/R/?func=dbin-jump-full&object_id=18670&local_base=GEN01-MCG02, accessed March 30, 2017.

Huang, Jin, Huug M. van den Dool, and Konstantine P. Georgarakos
1996 Analysis of Model-Calculated Soil Moisture over the United States (1931–1993) and Applications to Long-Range Temperature Forecasts. Journal of Climate 9(6): 1350–1362.

Kalnay, E., M. Kanamitsu, R. Kistler, et al.
1996 The NCEP/NCAR 40-Year Reanalysis Project. Bulletin of the American Meteorological Society 77(3): 437–471.

Montecinos, Aldo, and Patricio Aceituno
2003 Seasonality of the ENSO-Related Rainfall Variability in Central Chile and Associated Circulation Anomalies. Journal of Climate 16(2): 281–296.