water

MDPI

Article

# General Assessment of the Operational Utility of National Water Model Reservoir Inflows for the Bureau of Reclamation Facilities

**Francesca Viterbo [1,2,*], Laura Read [3], Kenneth Nowak [4], Andrew W. Wood [3], David Gochis [3], Robert Cifelli [2] and Mimi Hughes [2]**

1 CIRES, University of Colorado Boulder, Boulder, CO 80309, USA
2 NOAA/Physical Sciences Division, Boulder, CO 80305, USA; rob.cifelli@noaa.gov (R.C.); mimi.hughes@noaa.gov (M.H.)
3 National Center for Atmospheric Research, Boulder, CO 80301, USA; lread@ucar.edu (L.R.); andywood@ucar.edu (A.W.W.); gochis@ucar.edu (D.G.)
4 Research and Development, Bureau of Reclamation Denver, Denver, CO 80215, USA; knowak@usbr.gov
* Correspondence: francesca.viterbo@noaa.gov; Tel.: +1-303-497-6702

Received: 8 September 2020; Accepted: 14 October 2020; Published: 16 October 2020




**Abstract:** This work investigates the utility of the National Oceanic and Atmospheric Administration's National Water Model (NWM) for water management operations by assessing the total inflow into a select number of reservoirs across the Central and Western U.S. Total inflow is generally an unmeasured quantity, though critically important for anticipating both floods and shortages in supply over a short-term (hourly) to sub-seasonal (monthly) time horizon. The NWM offers such information at over 5000 reservoirs across the U.S., however, its skill at representing inflow processes is largely unknown. The goal of this work is to understand the drivers for both well performing and poor performing NWM inflows such that managers can get a sense of the capability of NWM to capture natural hydrologic processes and in some cases, the effects of upstream management. We analyzed the inflows for a subset of Bureau of Reclamation (BoR) reservoirs within the NWM over the long-term simulations (retrospectively, seven years) and for short, medium and long-range operational forecast cycles over a one-year period. We utilize ancillary reservoir characteristics (e.g., physical and operational) to explain variation in inflow performance across the selected reservoirs. In general, we find that NWM inflows in snow-driven basins outperform those in rain-driven, and that assimilated basin area, upstream management, and calibrated basin area all influence the NWM's ability to reproduce daily reservoir inflows. The final outcome of this work proposes a framework for how the NWM reservoir inflows can be useful for reservoir management, linking reservoir purposes with the forecast cycles and retrospective simulations.

**Keywords:** national water model; water management; total inflow; bureau of reclamation; reservoir characterization; reservoir inflows; retrospective evaluation; forecast evaluation; distributed hydrological modelling; reclamation information sharing environment (rise)

## 1. Introduction

Management decisions regarding reservoir storage and releases require accurate and timely knowledge of the contributing inflows from the basin [1–5]. Most global watersheds still remain completely ungauged, partially gauged or poorly gauged [6–8]. This lack of hydrological information also limits the possibility to measure reservoir inflows, largely due to high uncertainties in over-lake precipitation estimates and unmeasured streamflow from minor tributaries [9,10]. Even if some efforts have been made in using satellite data and radar data to implement reservoir inflow estimates [11–14]

there is still need to invest more money into field survey data and more ground truth measurements. Knowing these limitations, the field of hydrology has begun to shift more emphasis toward large scale hydrologic modelling [15–19], which can take advantage of remote sensing estimates for unmeasured quantities.

In the Western U.S., especially, relatively large contributions of drainage into a reservoir may be ungauged or seasonally gauged, leaving managers to rely on qualitative or climatological data which may not be robust for real-time operations [20]. At the same time, the adoption of forecasts into operations is limited due to uncertainty in the accuracy of inflow stream forecasts [21,22]. Since most reservoirs are multi-purpose, competing objectives such as flood control, irrigation, and instream flow requirements mean that management is dynamic and requires accurate information about the meteorological and hydrological conditions from the preceding days/months as well as some idea of the near-future forecast conditions.

Reservoir inflow forecast methods and applications are generally grouped as statistical, such as learning-based algorithmic, or utilize simulated streamflow from a process-based hydrologic model [23]. Approaches such as machine learning and neural networks that use training algorithms to represent system dynamics have been applied specifically to predict reservoir inflow [24,25], as have methods that use auto-regressive (integrated) moving average (ARMA/ARIMA) models [26]. Implementation of such models generally requires a rich dataset for which there are significant capital costs in data collection and model development per individual basin (or case study). Others have applied optimization principles to develop reservoir inflows forecasts, using linear programming, dynamic [27], and stochastic dynamic programming [28]. In these applications, inflows are usually stochastic variables drawn from historical data [29,30]. Previous studies have shown the potential benefits of including hydro-meteorological variables and inflow streamflow forecasts for short-term reservoir planning [1,31], though this is still not common practice among agencies. The purpose of this paper is to investigate the utility of a new operational forecasting system, the National Oceanic Atmospheric Administration's (NOAA's) National Water Model (NWM) in informing the inflows of reservoirs for Bureau of Reclamation (BoR) operations.

The BoR manages more than 600 dams and reservoirs over 17 Western States, bringing water to more than 31 million people, irrigating 10 millions of acres of lands, being the country's largest wholesaler of water and the country's second-largest producer of hydropower (after the US Army Corp of Engineers [32]). The BoR operates water projects across the Western U.S. These water infrastructure projects provide agricultural, municipal, and industrial water to about one-third of the population in the American West. Projects are operated pursuant to applicable treaties, compacts, federal and state regulations, contracts, and other requirements. Within this landscape, operational flexibilities are used to strike an equitable balance between the many, at times competing, services these projects provide (e.g., hydropower, recreational releases, etc.). To effectively deliver these services, accurate forecasts of the total inflow into BoR reservoirs is fundamental for managing water in a safe and sustainable way, through its storage and release into the river network.

The BoR widely depends on forecast products from Weather Forecast Offices (WFOs) and River Forecast Centers (RFCs) to inform different types of management decisions. For example, for spring runoff, basin simulations provide general snowpack and antecedent moisture conditions to help predict expected seasonal runoff. At shorter timescales, RFC streamflow and runoff forecasts based on operational meteorological products are used to inform release scheduling decisions, from flood control to instream flows. Despite the ability to transfer this water management information at local offices, the recent 2019 Federal Action plan [33] indicates the need to "improve and expand use of water forecast information to benefit water management outcomes". Example needs from findings in this report include: (i) improve consistency in hydrologic forecast products from days to seasons (subseasonal to seasonal forecasting) and (ii) incorporate key operational weather and climate outlooks, (iii) increase the reliability and skill of the currently available precipitation and hydrologic (e.g., streamflow) forecast products and (iv) incorporate them in water management decisions, (v) expand

the usage of continental scale hydrologic modeling (such as NWM) in order to expand geographic coverage of forecast products to ungauged areas and (vi) provide predictive capabilities on key water budgeting components such as snow states, soil moisture, evapotranspiration, runoff, and streamflow. While recent publications on NWM and operational interest concentrate streamflow evaluation [34], this paper focuses on understanding the ability of the NWM to capture total reservoir inflow, defined as the total contribution of tributaries (gauged and ungauged) into reservoirs, and the conditions for which the NWM can be useful for water use planning. The intent of this paper is to examine and explain the behavior of NWM reservoir inflows to inform BoR planning and operations, especially in reservoirs and watersheds that do not have total inflow forecast information available and for which managers are charged with making short to mid-order decisions (0−30 day) based on inflow information.

In a two-part approach, we first conduct a quantitative evaluation of NWM forecasts and multi-year retrospective simulations. Second, we use the assessment results to investigate the physical mechanisms that drive differences in performance across reservoirs and contextualize this performance within the BoR's reservoir operations objectives. As a final outcome of this work, we develop a set of recommendations for integrating the NWM into BoR reservoir operation decisions. This work is the first of its kind to evaluate a large set of NWM reservoir inflows for the specific purpose of exploring pathways for use in BoR operations and management. This paper is organized as follows: Section 2 describes the materials and methods for the inflow analyses, Section 3 illustrates the results, Section 4 presents a discussion of the results, and Section 5 offers concluding remarks.

## 2. Materials and Methods

### 2.1. NWM Background

Since inception into operations on the National Centers for Environmental Prediction (NCEP) high-performance computing system in August 2016, the NOAA−NWM has been providing real-time distributed continuous hydrologic forecasts across the Continental United States (CONUS), its contributing watersheds, and Hawaii, in four temporal cycles. The NWM development cycle is such that it is difficult to complete a detailed analysis of one version before the next version is updated and pushed for operations. During this study the NWMv1.2 upgraded to v2.0, so for completeness (availability of a full year of forecasts), we focus this study on v1.2. As of NWMv1.2, the cycles consist of an analysis and assimilation (hourly), short range (hourly from 1−18 h), medium range (six hourly, 1−10 days), and long range (daily, four-member ensemble, 1−30 days); for details on the cycles and configurations, see the service agreement [35], the NOAA official website [36] and references therein. The model base for NWM is the Weather Research Forecast (WRF)−Hydro system, which is described by [37]. The NWM configuration is an instance of the WRF−Hydro modeling framework that couples the Noah−MP land surface model at a one-kilometer grid spacing to an overland routing diffusive wave scheme at 250 m; the resulting infiltration excess is mapped from its catchments to a channel network based on the National Hydrography Dataset (NHD)Plus v2.1 [38].

The outputs of the NWM include streamflow and other hydrologic variables that describe hydrological states on the land surface (moisture and heat fluxes, snowpack characteristics, soil moisture and infiltration variables, etc.), at ~2.7 million river reaches and 5459 waterbodies (lakes and reservoirs as of NWM v2.0).

The physics of the WRF−Hydro channel and waterbody routing and representation are fully described in [37]. Briefly, the NWM collects streamflow from contributing streams and aggregates it within a waterbody to calculate the total inflow per time step. The outflow is then routed using a simple level-pool scheme, whereby outflow is controlled by static physical parameters (surface area and maximum depth) and parameterized outflow mechanisms (orifice and weir). Waterbodies in the NWM are required to have a minimum surface area of one $km^2$, contain at least one "on-channel" inflow tributary reach, and have exactly one discharge point. Note that due to data limitations, the NWM does not currently distinguish between a managed reservoir and a natural lake, though as data services

like the BoR Reclamation Information Sharing Environment (RISE) data service [39], the US Army Corps Water Management System (CWMS) and Access2Water (AWS) become more streamlined, it may be possible to ingest classification information to inform such parameters and discharge mechanisms. Development of data services to ingest real-time observations and forecasts of reservoir outflows is on-going and planned for implementation in a future NWM version. Kim et al. (2020) [40] explored the option to couple the NWM with a more sophisticated reservoir management model: the Hydrologic Evaluation Center's ResSim. In this configuration, NWM inflows were used as inputs into ResSim, which computed outflows based on rule curves and operating plans; results for the Russian River case showed significant improvement in representing managed outflows over the default NWM level pool scheme. This type of advancement provides a pathway for using NWM inflows to couple with more advanced outflow models, strengthening the motivation for understanding and improving NWM reservoir inflow processes.

### *2.2. Reconstructed Total Inflow Observations*

The inflow observations used for these comparisons are not direct observations, but are reconstructed inflow estimates based on reservoir daily release and change in reservoir storage. Reservoir evaporation and change in bank storage may be included in the calculations, depending on the availability of the required parameters per reservoir [41].

The BoR distributes this information through the Reclamation Information Sharing Environment (RISE) data portal, a publically available data service to view, access, and download historic and near real-time data [39].

Generally, the reconstructed regulated total inflow calculation attempts to reproduce the true total inflow of a reservoir, taking into account both natural and regulated components upstream. A second product, unregulated inflow, removes the contributions from upstream reservoir operations but still includes diversions/depletions to provide a naturalized inflow that is more independent of management [42]. In the present paper, we use the BoR regulated total inflow to compare with NWM total inflow variable, since our main target is to assess the NWM's ability to capture the observed states. Additionally, unregulated inflows were only available for three of the reservoirs in this selection, and after a preliminary analysis, these reconstructions were not directly applicable for the purposes of this project.

Finally, although the regulated total inflow is the best available observation for this analysis, it is a derived product and carries intrinsic uncertainties. These uncertainties and observation errors vary site by site, mainly impacted by the quality of elevation, evaporation and (in some cases) wind and bank storage measurements. For this reason, the quality of the reconstructed reservoir inflows in this study has been assessed through qualitative inspection of the hydrographs and through input from BoR. As a result, each reservoir in the analysis has been assigned a code to denote quality from red (noted data defects) to green (high fidelity), where orange is moderate fidelity.

Additionally, during periods of ice coverage or near-surface freezing, the BoR observations are not available, thus some time gaps exist in the observed inflows; reservoirs with more than 30% missing data over the total period were discarded.

### *2.3. Analysis Methods*

#### 2.3.1. Reservoir Selection

The reservoirs analyzed in this study were selected to include as many reservoirs as possible in multiple regions of interest to the BoR. The reservoirs selected met two primary criteria: their presence in both NWMv1.2 and NWMv2.0, and the availability and quality of observations from the RISE data service. The intersection of NWM and RISE reservoirs resulted in 42 reservoirs, which serve as the basis for this analysis. From these initial 42 reservoirs, we then discarded reservoirs with more than 30% missing data in either the retrospective or forecast period of interest, respectively, resulting in a

final count of 33 reservoirs for both retrospective and forecast evaluation, eight reservoirs only for the forecast analyses (since they had 30% missing data for the retrospective time-period but they matched the criteria for the forecast time-period) and one reservoir (Lake Berriessa and Monticello Dam, *ber*) that matched the data availability criteria for the retrospective, but not for the forecast evaluation. Figure 1 shows the locations of the reservoirs and associated watersheds colored by dominant Ecoregion (Level 3), which are groupings of geographic regions by similar physical attributes (see [43]).

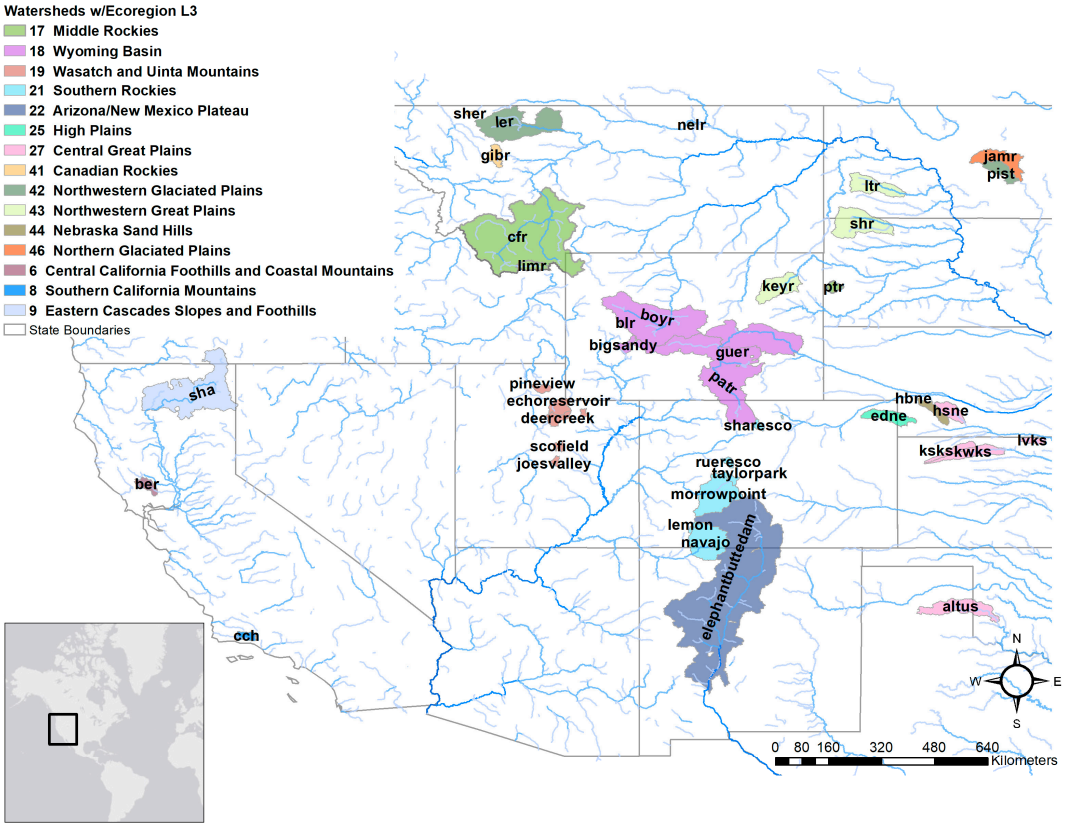



**Figure 1.** Map of watersheds for the 42 reservoirs included in this study; polygons are colored according to the level three Ecoregion with the dominant land area.

### 2.3.2. Reservoir Characterization

One goal of this effort is to identify the driving mechanisms for NWM inflow performance so that the results can be generalized and the model may be improved. Thus in the evaluation, we classified reservoirs according to characteristics that may influence the model performance: Ecoregion, percent of watershed area calibrated, drainage area of the contributing watershed, percent assimilated area upstream (described in this section and in Section 3.2 below), number of lakes upstream and percentage of snow contribution over total precipitation. Some of the characteristics mentioned are referring to physical characteristics of the watershed itself (e.g., Ecoregion, drainage area, percentage of snow contribution), while others are particularly linked to the NWM architecture (percent of watershed area calibration, percent of assimilated area upstream). Others, like Ecoregion, link both physical characteristics of the watersheds and the NWM architecture, since the NWM parameters are calibrated in representative watersheds and then regionalized as a function of Ecoregion in v1.2. For further information about NWM calibration, the reader is referred to Gochis et al., 2019 [44].

The characteristics were attributed as follows: Ecoregions—Level 3; the drainage area of the reservoir's contributing watershed was calculated using the NWM flow network; the percentage of snow contribution was attributed using the GAGES2 dataset [45] where gauges from this dataset were associated with the watersheds of each reservoir. The variable: "snow percent of total precipitation estimate, mean for period 1901−2000" was used to define this characteristic, following McCabe and Wolock (2010) [46]. The percent calibrated area is the NWMv1.2 calibrated area within the reservoir watershed divided by the total contributing drainage area. The percent upstream assimilated area (or, percent gauged area) is calculated by summing the contributing areas that belong to assimilation gauges in the NWM and dividing by the total reservoir drainage area. This characteristic indicates both how much assimilation correction has been performed upstream of the reservoir for the NWM forecasts, and how many streamflow stations are present upstream of each reservoir. If the percent is equal to zero, it means not only that the NWM forecast is unadjusted by assimilation, but it also means that the area upstream of the reservoir is ungauged by United States Geological Survey (USGS). Finally, the number of lakes upstream is defined as the number of NWM lakes or reservoirs contributing to the inflow of a given reservoir, identified by tracing upstream on the flow network. A summary of the reservoirs and characterizations are given in Table 1.

In addition to these physical and model-based classifications, we also used information from the National Inventory of Dams (NID) database [47] to assign a primary purpose to each reservoir. Using expert knowledge from BoR partners and local offices, the primary purpose was matched with the NWM cycles that are most applicable for each reservoir purpose (Table 2). This association was intended to emphasize the relevance of the NWM for specific reservoir purposes. The results from the responses highlighted that, for example, the NWM long range forecast is more relevant for irrigation, while the short and medium range forecast is more appropriate for flood control operations. A summary of the stated primary purpose for each reservoir according to NID is also given in Table 1, last column.

#### 2.3.3. NWM Inflow Assessment

The inflow evaluation was conducted on the 34 selected reservoirs through two sets of model simulations: (1) "retrospective": a seven-year, hourly, open-loop (no assimilation) retrospective simulation using the NWMv1.2 and NWMv2.0 forced by the National Land Data Assimilation System (NLDAS−2) [48] downscaled and regridded on the 1-km NWM land-surface grid, and (2) "forecast": a one-year simulation of the operational NWMv1.2 short- (0−18h), medium- (0−10 days) and long-range (0−30 days) forecast configurations, forced with NWM operational forcings, and with data assimilation active. Model configuration details, including sources for meteorological forcings, and information on downloading model outputs are available at Office of Water Prediction (2016) and references therein [36].

Assessing both the forecast and retrospective simulations is important for obtaining a robust evaluation and providing a more complete picture of overall NWM performance. The performance of the NWM in forecast mode incorporates meteorological uncertainties coming from the weather forecast forcings, and indicates the model's general ability to capture shorter events across lead times and to provide utility for operational decisions. In comparison, the retrospective is forced by weather variables coming from analyses (NLDAS−2), and provides insight on overall model physics and representation of seasonal processes that can be informative for longer-term planning. In dissecting where and when the model shows congruence between the forecast and retrospective performance, we can draw conclusions about its utility for reservoir management.

**Table 1.** Reservoir and drainage basin characteristics. The numbers in the Ecoregion column correspond with the legend of Figure 1. In the main purpose column abbreviations are used as following: I for Irrigation, FC for Flood Control, H for Hydroelectric, FW for Fish and Wildlife pond. BoR regions are abbreviated as MP for Mid Pacific, GP for Great Plains, UC for Upper Colorado. Colors in column one indicates quality from reconstructed total inflows, from red (noted data defects) to green (high fidelity), where orange is moderate fidelity.

| Reservoir | Full Name | State | BoR Region | Ecoregion | Snow over total Precip (%) | N. of Lakes | Gauged Area (%) | Area Lake ($km^2$) | Area Basin ($km^2$) | Calib. Area (%) | Main Purpose |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ber | Lake Berryessa And Monticello Dam | CA | MP | 6 | 2 | 0 | 20 | 76 | 1469 | 0 | I |
| cch | Cachuma Lake And Bradbury Dam | CA | MP | 8 | 5 | 1 | 84 | 13 | 1081 | 18 | I |
| sha | Shasta Lake | CA | MP | 9 | 33 | 20 | 53 | 113 | 21240 | 72 | FC |
| limr | Lima Rsv Upstr of Clark Canyon Rsv | MT | GP | 17 | 47 | 4 | 7 | 18 | 1463 | 0 | I |
| ptr | Pactola Rsv nr Rapid City | SD | GP | 17 | 47 | 1 | 93 | 3 | 820 | 25 | FC |
| cfr | Canyon Ferry Lake nr Helena | MT | GP | 17 | 38 | 15 | 14 | 136 | 41218 | 8 | FC |
| blr | Bull Lake Rsv on Bull Lake Creek | WY | GP | 17 | 62 | 0 | 91 | 12 | 535 | 90 | FC |
| bigsandy | Big Sandy Rsv | WY | UC | 18 | 47 | 0 | 83 | 9 | 1016 | 0 | I |
| boyr | Boysen Rsv | WY | GP | 18 | 31 | 7 | 61 | 78 | 19952 | 7 | FC |
| patr | Pathfinder Rsv | WY | GP | 18 | 32 | 6 | 81 | 89 | 27583 | 2 | H |
| guer | Guernsey Rsv | WY | GP | 18 | NA | 9 | 94 | 8 | 40995 | 1 | H |
| joesvalley | Joes Valley Rsv | UT | UC | 19 | 47 | 0 | 0 | 5 | 381 | 0 | FC |
| moonlake | Moon Lake Rsv | UT | UC | 19 | 63 | 0 | 70 | 3 | 290 | 0 | I |
| scofield | Scofield RSV | UT | UC | 19 | 58 | 0 | 63 | 11 | 402 | 17 | I |
| pineview | Pineview Rsv | UT | UC | 19 | 55 | 0 | 45 | 10 | 792 | 0 | I |
| deercreek | Deer Creek Rsv | UT | UC | 19 | 52 | 0 | 81 | 10 | 1419 | 0 | I |
| echoreservoir | Echo Rsv | UT | UC | 19 | 51 | 1 | 92 | 5 | 1874 | 0 | FC |
| rockportreservoir | Rockport Rsv | UT | UC | 19 | 56 | 0 | 87 | 4 | 866 | 0 | I |
| lemon | Lemon Rsv | CO | UC | 21 | 55 | 0 | 0 | 3 | 174 | 0 | FW |
| navajo | Navajo Rsv | NM | UC | 21 | 38 | 3 | 76 | 62 | 8302 | 2 | FC |
| morrowpoint | Morrow Point Rsv | CO | UC | 21 | NA | 3 | 66 | 4 | 9420 | 42 | FC |
| rueresco | Ruedi Rsv nr Basalt | CO | GP | 21 | 61 | 0 | 21 | 4 | 579 | 0 | H |
| taylorpark | Taylor Park Rsv | CO | UC | 21 | 62 | 0 | 50 | 8 | 658 | 50 | I |
| elephantbuttedam | Elephant Butte Rsv | NM | UC | 22 | NA | 15 | 81 | 23 | 77826 | 7 | FC |
| lvks | Lovewell Dam | KS | GP | 27 | 9 | 0 | 66 | 12 | 896 | 65 | FC |
| hsne | Harry Strunk Lake (Medicine Creek Dam) | NE | GP | 27 | 9 | 0 | 0 | 7 | 2295 | 0 | FC |
| gibr | Gibson Rsv | MT | GP | 41 | 55 | 0 | 92 | 5 | 1436 | 0 | I |
| sher | Lake Sherburne | MT | GP | 41 | 68 | 0 | 47 | 5 | 171 | 23 | I |
| ler | Lale Elwell Spillway Transducer | MT | GP | 42 | 24 | 10 | 62 | 60 | 11378 | 0 | FC |
| pist | Pipestem Dam | ND | GP | 42 | 17 | 0 | 68 | 8 | 2342 | 0 | FC |
| shr | Shadehill Rsv on the Grand River | SD | GP | 42 | 14 | 1 | 83 | 21 | 7751 | 0 | FC |
| ltr | Lake Tshida (Heart Butte) | ND | GP | 43 | 17 | 1 | 89 | 13 | 4438 | 9 | FC |
| hbne | Hugh Butler Lake (Red Willow Dam) | NE | GP | 44 | 10 | 0 | 0 | 6 | 1775 | 0 | FC |
| jamr | Jamestown Rsv | ND | GP | 46 | 18 | 6 | 65 | 18 | 4132 | 66 | FC |
| kwks | Kirwin Rsv at Kirwin | KS | GP | 27 | 0 | 0 | 92 | 20 | 3634 | 0 | FC |
| altus | Altus Dam | OK | GP | 27 | 94 | 0 | 94 | 18 | 6637 | 94 | I |
| sharesco | Shadow Mountain Rsv | CO | GP | 21 | 0 | 0 | 77 | 7 | 477 | 0 | I |
| edne | Enders Dam and Dike | NE | GP | 25 | 0 | 0 | 0 | 7 | 3310 | 0 | FC |
| htoothr | Horsetooth Rsv nr Fort Collins | CO | GP | 21 | 0 | 0 | 0 | 7 | 42 | 0 | I |
| keyr | Keyhole Rsv | WY | GP | 43 | 1 | 0 | 86 | 3 | 5054 | 1 | FC |
| ksks | Keith Sebelius Lake (Norton Dam) | KS | GP | 27 | 84 | 0 | 85 | 8 | 1799 | 84 | FC |
| nelr | Nelson rsv at Dam 10 miles NW of Saco | MT | GP | 42 | 0 | 0 | 0 | 17 | 81 | 0 | I |

**Table 2.** Associations of National Water Model (NWM) model forecast cycles with reservoir purposes.

| | Model Cycle | | |
|---|---|---|---|
| **Purpose** [1] | **Short Range** | **Medium Range** | **Long Range** |
| Water Supply | | | x |
| Irrigation (12) | x | | x |
| Hydropower (3) | x | x | x |
| Flood Control (18) | x | | x |
| Ecosystem (1) | | x | x |

[1] numbers in the parentheses denotes the reservoir count in this study; "x" indicates that the cycle is relevant for the project purpose; dark grey = highly relevant, light grey = potentially relevant.

The total reservoir hourly inflow from the NWM was aggregated to a daily timestep for comparison with the BoR reconstructed daily inflow observations as follows: hours with same lead times (e.g., six-hour lead time) with valid times on each day were averaged for each day to create aggregated daily forecasts. This was done for each forecast cycle (i.e., short-, medium-, and long-range) at their respective output frequencies (i.e., hourly, three-hour, and six-hour, respectively).

NWM skill was assessed using three common goodness of fit statistics widely utilized in hydrology: Pearson's correlation coefficient ($r$), Nash–Sutcliffe Efficiency ($NSE$), and relative bias ($relBias$) [49–51]. We use $relBias$, which is a bias normalized by average inflow (Equation (1)),

$$relBias = \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n} O_i} \quad (1)$$

where $O_i$ is the observation value and $P_i$ is the forecast value, in order to compare reservoirs of different sizes with one another. These skill scores were selected to give an indication of response congruence ($r$), volumes ($relBias$), and an overall measure that considers both timing and volume ($NSE$).

## 3. Results

### 3.1. Retrospective Analyses

Retrospective NWM total inflow was analyzed in order to provide insight on overall model physics and representation of seasonal processes. Since NWM was in the process of transitioning from v1.2 to v2.0, we did an initial check on retrospective performance differences between the two versions to ensure analysis from v1.2 is still relevant. Goodness of fit metrics were computed on daily streamflow using the retrospective model simulations from NWMv1.2 and v2.0 against BoR observations. Based on skill scores for $r$, $relBias$, and $NSE$, we note that inflows were generally improved from v1.2 to v2.0 across most reservoirs, and that reservoirs with low skill in v1.2 also generally had low skill in v2.0, indicating a persistent issue at these locations. Because a sufficient time period of forecasts were only obtainable from NWMv1.2, and the retrospective evaluation confirms a similar pattern between v2.0 and v1.2, NWMv1.2 is the focus of this analysis. Figure 2 shows the goodness of fit scores for $r$, $relBias$, and $NSE$ (top to bottom) for the retrospective evaluation in v1.2 (dark) and v2.0 (light), where the $x$-axis is sorted from left to right by an increasing number of lakes upstream (with the count in brackets). The colors from green to red indicate the quality of the observations according to the description in Section 2.2.

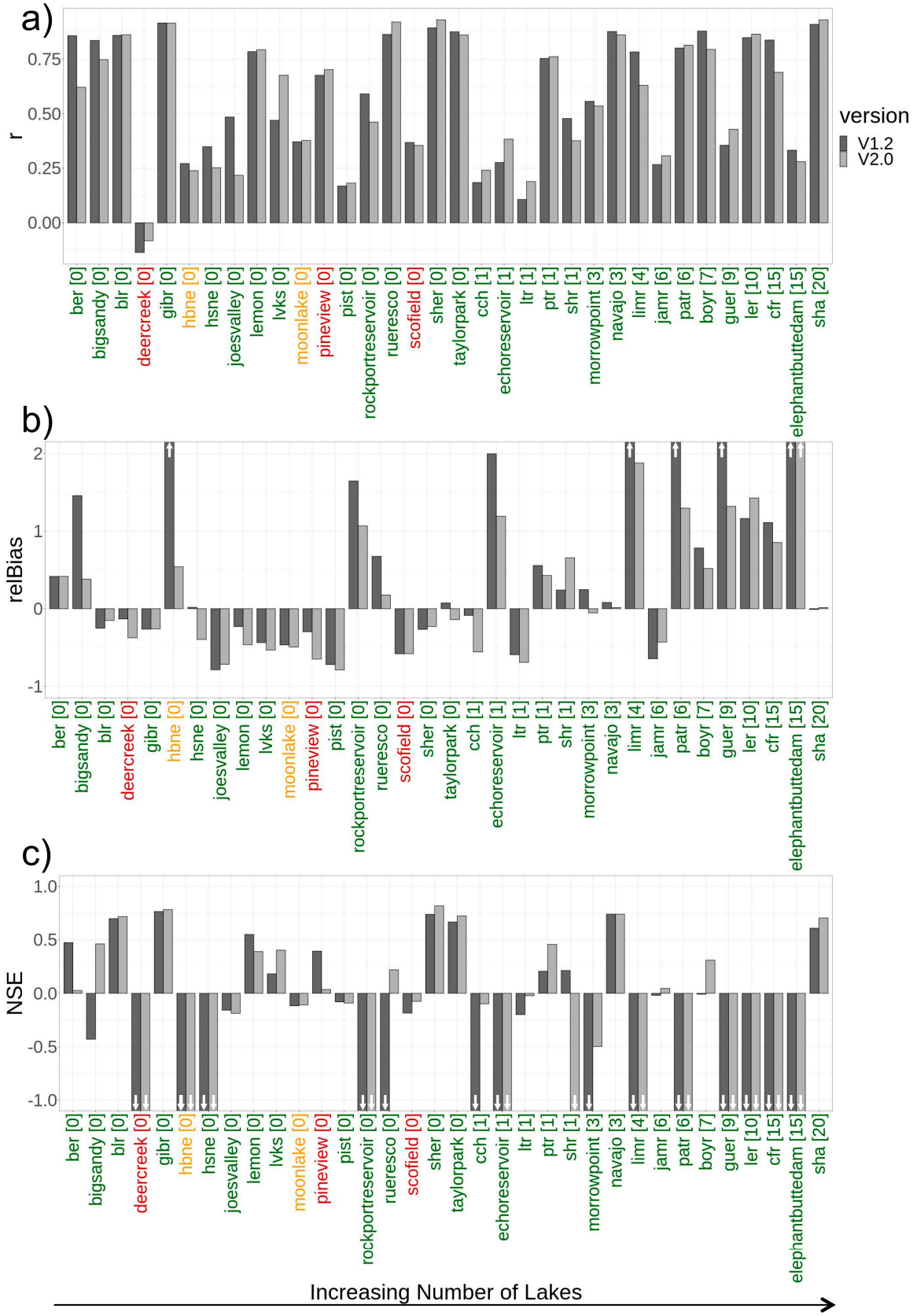



**Figure 2.** Goodness of fit scores from the NWM v1.2 (dark grey) and NWM v2.0 (light grey) retrospective total inflow evaluation, ordered by increasing number of lakes. From top to bottom: *r* (**a**), *relBias* (**b**) and *NSE* (**c**).

Since the retrospective runs in 'open-loop', it does not account for managed upstream flows or active data assimilation. While we would expect the NWM to be able to simulate inflows into reservoirs with no management upstream better than those with multiple upstream lakes (Figure 2), the results are mixed. Nevertheless, the high positive *relBias* in managed watersheds indicates that NWM is simulating the unmanaged hydrology, which is not able to account for engineered inflow reductions from upstream reservoir operations.

In parsing the NWM inflow performance by the full set of reservoir characteristics (as listed in Table 1), expected patterns with percent calibrated area (Figure A1) and drainage basin size (Figure A2) emerged. In reservoirs with a higher percent of calibrated area, NWM inflows show better performance, especially for lakes with calibrated areas greater than ~10%. The relationship between modeled inflow performance and lake drainage basin size is inverse, such that the model is able to simulate inflows in lakes with smaller basins better than larger ones, likely because larger basins contain more upstream management (e.g., diversions, lakes, etc.). Finally, snow percentage contribution over total precipitation revealed better retrospective statistical scores for reservoirs with higher snow percentage contribution (Figure A3). See additional material in Appendix A, Figures A1–A3 for visual reference to these findings.

### 3.2. Forecast Assessment

This section presents the analysis of the NWM v1.2 daily inflow forecasts from September 2017 to August 2018 for the short, medium and long range forecast cycles. The results are sorted by characteristic (number of lakes upstream, percentage of assimilated area upstream, solid precipitation contribution and retrospective ranking) in order to investigate the influence of upstream management and assimilation on the forecast performance, focusing also on the differences between snow-driven and rain-driven total inflow contributions and how these different hydrological regimes are handled by the NWM forecast. Sorting the forecast results by different characteristics helped to infer the impact of each of these different elements on model performance.

As in the retrospective evaluation, the forecast analysis indicates that upstream management is a factor in inflow performance (Figure 3). We note that the *r* in reservoirs with upstream management is generally good (>0.7), but the *relBias* is high, whereas in the lakes without upstream management (24 of the 36), the *relBias* indicates underestimation. This pattern is true across all forecast cycles.

Figure 4 illustrates how the percentage of upstream assimilated area, a proxy for gauged area, influences each forecast cycle and lead times within. To explain the mechanisms governing this behavior, a brief description of the NWM streamflow data assimilation follows. In the AnA cycle, which provides the forecast initial conditions, nudging data assimilation is applied to the streamflow state at ~7000 USGS streamflow gauge locations through a real-time data feed on the operational system. Nudging is also applied in the forecast but only observations near the initialization time are used and the effect of the nudge is relaxed (with two hour e-folding) as the forecast advances into the future, beyond observations available near the initial time. Thus, we expect to see evidence of this relaxation across the first 0–3 h in the short-range at many gages (top row, Figure 4). Lakes that have a small percent of assimilation upstream exhibit two behaviors: low skill from early lead times, or good performance through the entire cycle. In highly assimilated areas, on the contrary, the short-range in some cases manages to keep memory of assimilation, improving the results even past the e-fold time (e.g., ~5-h lead time or more for Pactola reservoir (*ptr)*, Echo reservoir (*echoreservoir)*, Elephant Butte Dam (*elephantbuttedam)*, Lovewell Dam (*lvks)*, etc.). In contrast, the effect of large assimilation on medium and long range tends to dissipate relatively quickly over the longer lead times. In general, the medium and long-range inflow forecasts are less reliable at longer lead times than the short-range.

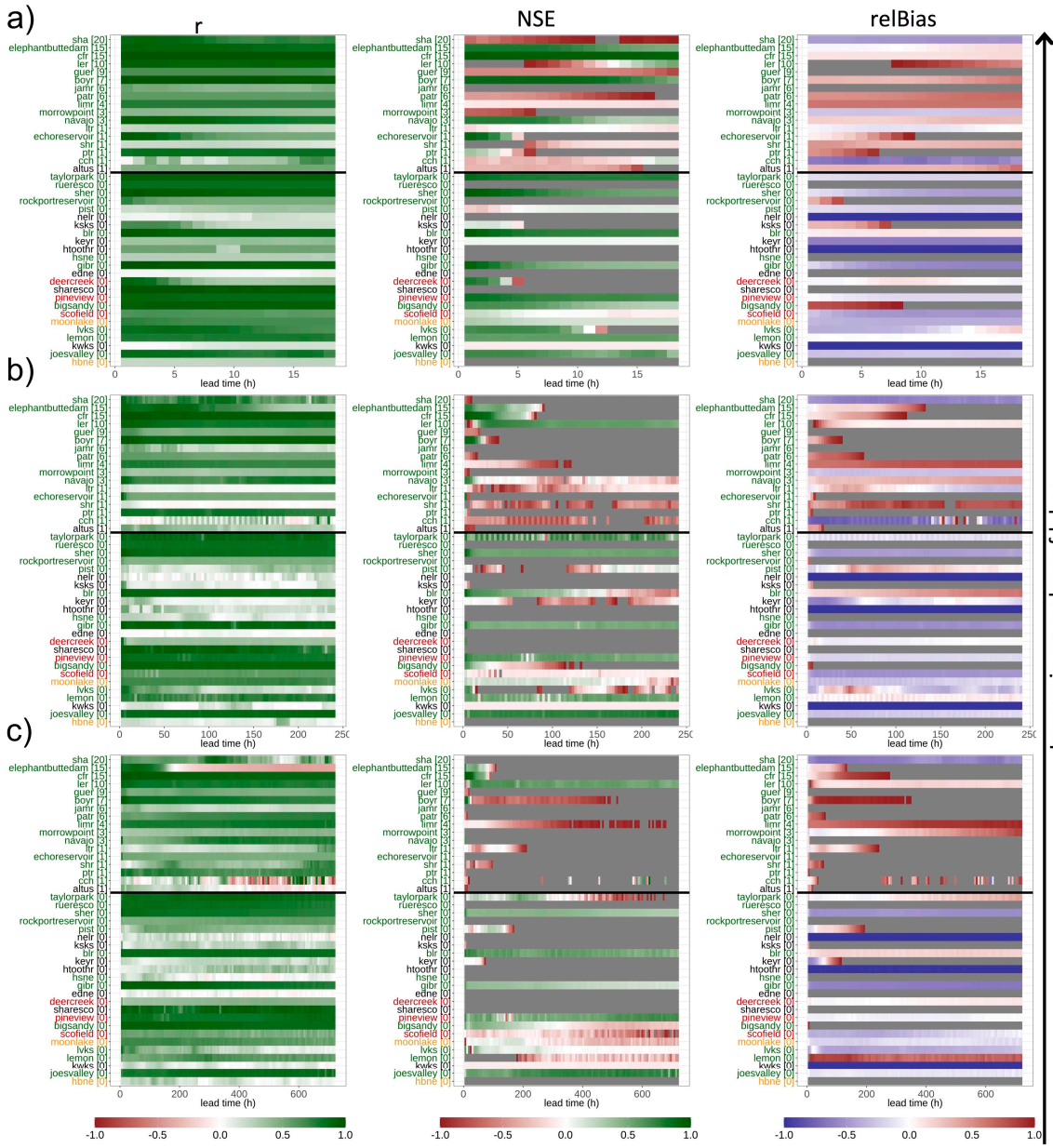



**Figure 3.** Performance of NWM reservoir inflows across lead times for one year, where reservoirs in each subplot are ordered from bottom to top by the number of upstream lakes (counts are listed in parenthesis on the *y*-axis with reservoir name). The black bold line separates reservoirs with more than zero lakes upstreams to the ones with zero lakes upstream.Each column from left to right columns illustrate *r*, *NSE*, *relBias*, where the metric is scaled by color density and grey is out of valid range (*relBias* > +1 or *relBias* < −1 and *NSE* < −1). For *r* and *NSE*, green values are higher performance and red is worse; and for *relBias*, blue is high-biased and red is low-biased. The subplot rows are forecast cycles: short (**a**), medium (**b**), long (**c**).

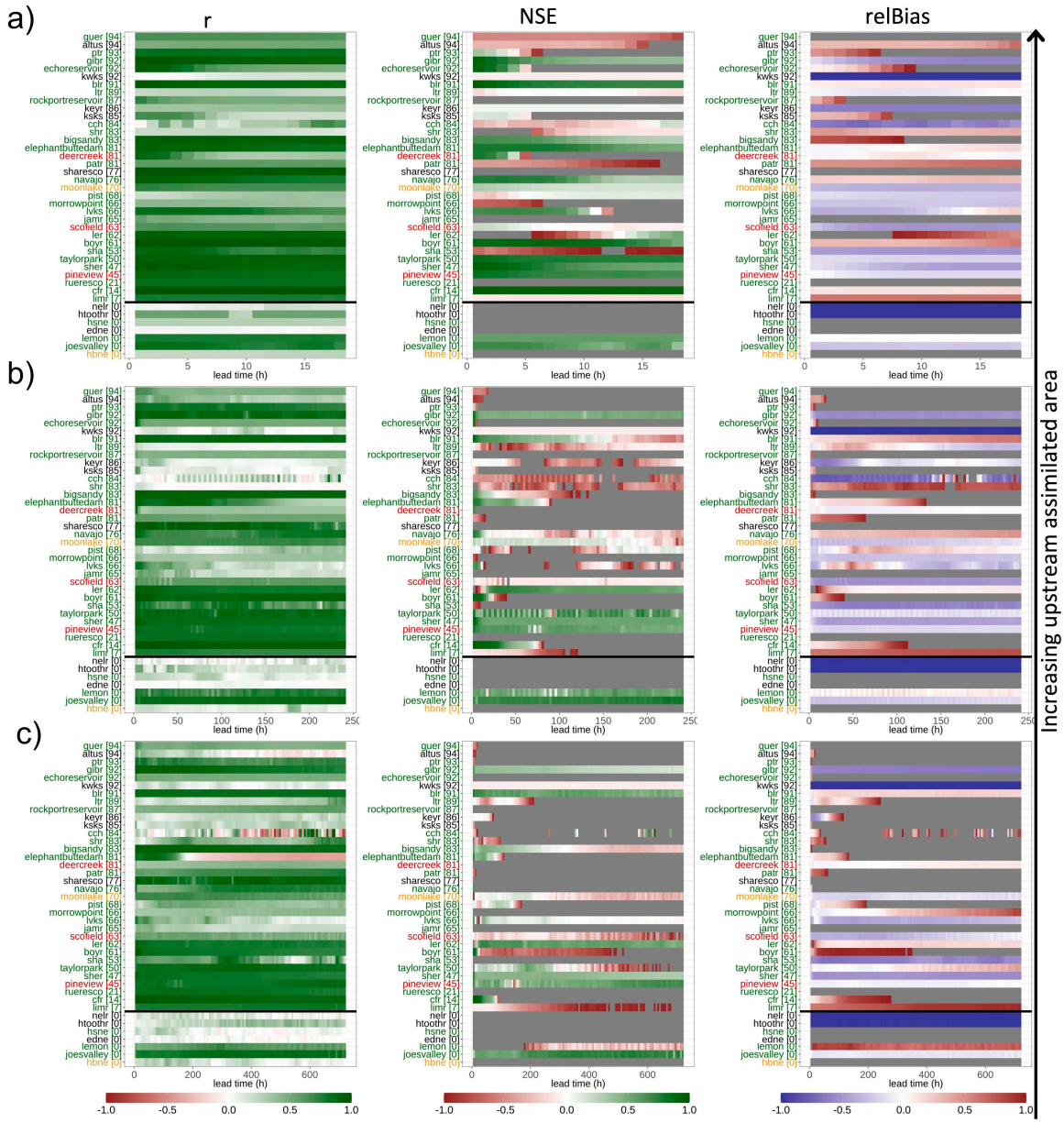



**Figure 4.** Performance of NWM reservoir inflows across lead times for one year, where reservoirs in each subplot are ordered from bottom to top by the percent of upstream assimilated area (percentages are listed in parenthesis on the *y*-axis with reservoir name). The black bold line separates reservoirs with no assimilation upstream to the ones that have assimilation upstream. Each column from left to right columns illustrate *r*, *NSE*, *relBias*, where the metric is scaled by color density and grey is out of valid range (*relBias* $> +1$ or *relBias* $< -1$ and *NSE* $< -1$). For *r* and *NSE*, green values are higher performance and red is worse; and for *relBias*, blue is high-biased and red is low-biased. The subplot rows are forecast cycles: short (**a**), medium (**b**), long (**c**).

In watersheds that are dominated by snow-driven processes, the *r* and *NSE* skill scores over the forecast cycles and across lead times tend to outperform those in rain-driven watersheds (Figure 5). In addition, snow-driven watersheds have also less degradation of the performance as a function of lead time in every forecast range (short, medium and long), even if some exceptions can still be found in the pattern. Rain-driven watersheds have the tendency to overestimate the volumes, according to *relBias* results, while reservoirs with greater snow contributions show more of a tendency to underestimate.

These findings are encouraging and indicate that the NWM can reasonably reproduce snow-melt runoff, an essential process for water resources management.

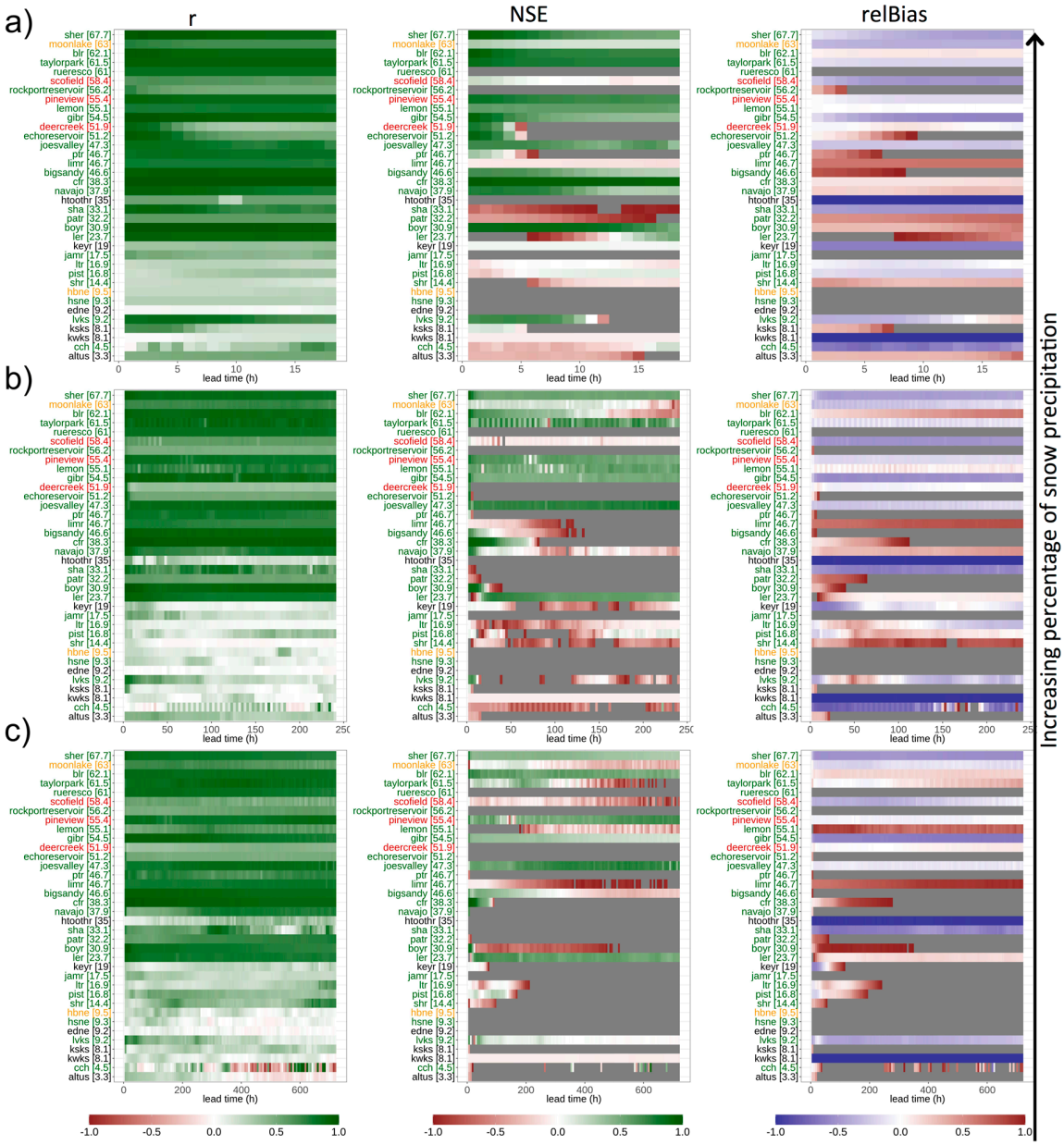



**Figure 5.** Performance of NWM reservoir inflows across lead times for one year, where reservoirs in each subplot are ordered from bottom to top by the percent contribution of snow in total precipitation (percentages are listed in parenthesis on the *y*-axis with reservoir name). Each column from left to right columns illustrate *r*, *NSE*, *relBias*, where the metric is scaled by color density and grey is out of valid range (*relBias* > +1 or *relBias* < −1 and *NSE* < −1). For *r* and *NSE*, green values are higher performance and red is worse; and for *relBias*, blue is high-biased and red is low-biased. The subplot rows are forecast cycles: short (**a**), medium (**b**), long (**c**).

From the retrospective assessment, we can confirm that snow processes are a significant driver in understanding inflow performance. For example, if we compare two reservoirs of similar size, gauged area, but in different regions, Lake Sherburne (*sher*, 67.7% snow-driven, Canadian Rockies ecoregion, 23% calibrated) and Lovewell Dam (*lvks*, 9.2% snow-driven, Central Great Plains ecoregion, 65% calibrated), the impact on the seasonal cycle is evident (Figure 6 and Table 3). The NWM is able to

replicate the seasonal and annual cycles in the snow-driven watershed contributing to Lake Sherburne reservoir (*sher*), while the peaks at Lovewell Dam (*lvks*) were not well aligned with the observations.

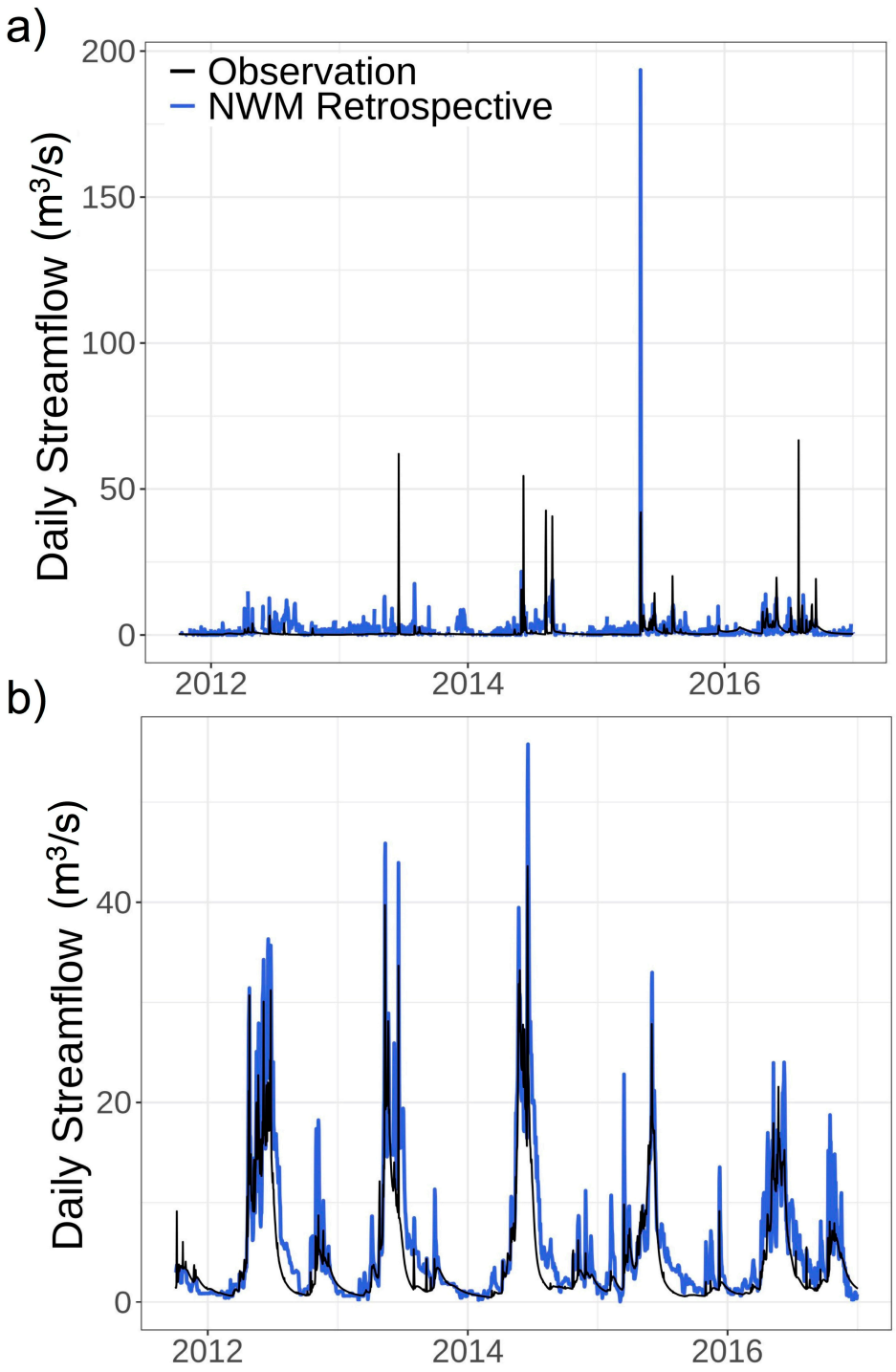



**Figure 6.** Retrospective comparison of inflows from two sites, Lake Sherburne (**a**) (*sher*, 67.7% snow-driven, Canadian Rockies, 23% calibrated) and Lovewell Dam (**b**) (*lvks*, 9.2% snow-driven, Central Great Plains, 65% calibrated). Observed is in black, model simulation in gray.

**Table 3.** Seasonal statistics for Lake Sherburne (sher) (panel a) and Lovewell Dam (lvks) (panel b). Rows represent statistical scores (from top to bottom): *r*, *relBias* and *NSE*. The columns represent the scores over each season: "total" refers to the total length of the retrospective time series, "ONDJFM" refers to the time-period from October to March, "AMJ" refers to the time-period from April to June, "JAS" refers to the time-period from July to September. The percentages in brackets next to each column represent the percentage observational data availability over the time period of interest.

| **(a)** | | | | |
|---|---|---|---|---|
| | **Total (86%)** | **ONDJFM (84%)** | **AMJ (87%)** | **JAS (91%)** |
| *r* | 0.47 | −0.12 | 0.64 | 0.25 |
| *relBias* | −0.44 | −0.6 | −0.33 | −0.42 |
| *NSE* | 0.18 | −0.47 | 0.35 | −2,21 |
| **(b)** | | | | |
| | **Total (86%)** | **ONDJFM (84%)** | **AMJ (87%)** | **JAS (91%)** |
| *r* | 0.89 | 0.73 | 0.87 | 0.92 |
| *relBias* | −0.27 | −0.24 | −0.15 | −0.58 |
| *NSE* | 0.74 | 0.44 | 0.71 | 0.12 |

### *3.3. Combined Assessment*

In an effort to synthesize the factors that influence inflow performance, Table 4 summarizes the reservoirs, sorted by Ecoregion, with the characteristics presented in Tables 1 and 2, and with the retrospective *NSE* and the forecast-cycle *NSE* scores at three lead times.

To summarize the results from the retrospective and forecast evaluation, Table 4 and Figure 7 offer visualizations of the combined analysis. Figure 7 provides a visual for assessing congruence between the retrospective and forecast by comparing the NWM inflow performance between the two configurations (ordered by *NSE* from bottom to top (best to worst) according to their retrospective ranking).

Taken together, the retrospective and forecast evaluation provide a comprehensive understanding of the overall reservoir performance: the retrospective evaluation consists of a multi-year time series and is forced by NLDAS analyses, that, even if have their own uncertainties, have assimilated observations in the meteorological forcings. On the contrary, the forecast runs span shorter time periods (with varying lead times for short, medium long range) and are forced by meteorological products that are derived from forecast models (High-Resolution Rapid Refresh (HRRR) for the short range, Global Forecast System (GFS) for the medium range, Climate Forecast System (CFS) for the long range). When the retrospective shows favorable performance, it is plausible to assume that in that case the NWM is able to sufficiently reproduce the physical processes over multiple years and seasons. The value of this analysis is in understanding variance in model performance by season and by different model version (since the meteorological forcings are the same) and using this knowledge to guide future development. In the cases with poor retrospective performance, we expect similar behavior in the forecast, unless a significant contribution from assimilation occurs at that site. Finally, when the retrospective skill score are unsatisfactory, but the forecast performance is good, it is difficult to infer conclusions but further investigations can be done in assessing the quality of the meteorological forcings (since, it is well known that the hydrological model cannot correct the errors coming from the forcings). This is the general set of guidelines to follow when synthesizing the two different evaluations (retrospective and forecast). Nevertheless, due to the NWM complexity and the propagation of uncertainties from the atmospheric forcings to the hydrology, the real answers are much more complicated and would require a detailed study for each single reservoir case. This focus is outside the scope of the present study, but hopefully this broader analysis will encourage researchers and stakeholders to deepen the analyses on the local level.

**Table 4.** Summary of reservoir characteristics and *NSE* skill scores for retrospective and forecasts (following page). The color densities indicate higher values for the characteristics (e.g., larger lake surface area) while the red to green densities for the *NSE* scores indicate the relative performance (red is worse and green is better). Ber does not show scores for the forecast since it had more than 30% missing data for the forecast period, as explained in Section 2.3.1. NA values for snow % are associated with places in which the nearest USGS was not included in the GAGES2 database.

| Reservoir | Main Purpose | Ecoregion | Snow over Total Precip [%] | N. of Lakes | Gauged Area [%] | Area Lake [km$^2$] | Area Basin [km$^2$] | Calib. Area [%] | NSE Retro | NSE Short Range | | | NSE Medium Range | | | NSE Long Range | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | +1 | +9 | +18 | +1 | +120 | +240 | +6 | +360 | +720 |
| ber | I | 6 | 2 | 0 | 20 | 76 | 1469 | 0 | 0.47 | | | | | | | | | |
| cch | I | 8 | 5 | 1 | 84 | 13 | 1081 | 18 | −16.61 | −0.35 | −0.34 | 0.18 | −0.40 | −0.54 | −0.52 | −0.28 | 0.01 | −3.17 |
| mltr | FC | 9 | 33 | 20 | 53 | 113 | 21,240 | 72 | 0.61 | −0.67 | −0.94 | −0.99 | −0.85 | −1.43 | −1.70 | −1.16 | −1.80 | −4.78 |
| cfr | FC | 17 | 38 | 15 | 14 | 136 | 41,218 | 8 | −8.58 | 0.95 | 0.95 | 0.91 | 0.95 | −5.39 | −8.04 | 0.96 | −4.63 | −4.49 |
| limr | I | 17 | 47 | 4 | 7 | 18 | 1463 | 0 | −9.45 | −0.08 | −0.11 | −0.14 | −0.11 | −0.95 | −1.42 | −0.09 | −0.98 | −1.38 |
| ptr | FC | 17 | 47 | 1 | 93 | 3 | 820 | 25 | 0.21 | 0.34 | −2.45 | −4.18 | 0.24 | −6.02 | −11.12 | −0.59 | −6.17 | −4.54 |
| blr | FC | 17 | 62 | 0 | 91 | 12 | 535 | 90 | 0.70 | 0.95 | 0.73 | 0.72 | 0.90 | 0.25 | −0.59 | 0.79 | 0.56 | 0.41 |
| boyr | FC | 18 | 31 | 7 | 61 | 78 | 19,952 | 7 | −0.01 | 0.89 | 0.89 | 0.50 | 0.89 | −1.65 | −3.28 | 0.87 | −0.77 | −1.19 |
| patr | I | 18 | 32 | 6 | 81 | 89 | 27,583 | 2 | −18.19 | −0.44 | −0.63 | −1.01 | −0.45 | −8.97 | −12.16 | −0.53 | −11.55 | −20.89 |
| bigsandy | I | 18 | 47 | 0 | 83 | 9 | 1016 | 0 | −0.43 | 0.66 | 0.48 | 0.27 | 0.65 | −0.93 | −2.18 | 0.56 | −0.09 | −0.22 |
| guer | H | 18 | NA | 9 | 94 | 8 | 40,995 | 1 | −9.29 | −0.54 | −0.54 | −0.82 | −0.54 | −4.02 | −13.00 | −0.54 | −15.95 | −25.41 |
| joesvalley | FC | 19 | 47 | 0 | 0 | 5 | 381 | 0 | −0.16 | 0.69 | 0.57 | 0.37 | 0.70 | 0.73 | −0.07 | 0.24 | 0.73 | 0.72 |
| echoreservoir | FC | 19 | 51 | 1 | 92 | 5 | 1874 | 0 | −39.71 | 0.90 | −12.24 | −24.13 | 0.81 | −36.08 | −41.84 | −1.17 | −35.89 | −34.95 |
| deercreek | I | 19 | 52 | 0 | 81 | 10 | 1419 | 0 | −7.35 | 0.78 | −5.03 | −2.52 | 0.67 | −3.62 | −4.89 | −1.27 | −2.73 | −2.50 |
| pineview | I | 19 | 55 | 0 | 45 | 10 | 792 | 0 | 0.39 | 0.83 | 0.70 | 0.66 | 0.79 | 0.58 | 0.24 | 0.72 | 0.51 | 0.68 |
| rockportreservoir | I | 19 | 56 | 0 | 87 | 4 | 866 | 0 | −16.16 | −1.36 | −86.33 | −95.40 | −11.29 | −158.93 | −191.73 | −56.05 | −157.75 | −141.63 |
| scofield | I | 19 | 58 | 0 | 63 | 11 | 402 | 17 | −0.19 | 0.22 | −0.02 | −0.08 | 0.18 | −0.06 | −0.07 | 0.03 | −0.27 | −0.74 |
| moonlake | I | 19 | 63 | 0 | 70 | 3 | 290 | 0 | −0.12 | 0.31 | 0.17 | 0.17 | 0.28 | 0.07 | −0.56 | −0.95 | −0.21 | −0.32 |
| navajo | FC | 21 | 38 | 3 | 76 | 62 | 8302 | 2 | 0.74 | 0.75 | 0.58 | 0.28 | 0.75 | −0.29 | −0.32 | 0.65 | −9.28 | −10.82 |
| lemon | FW | 21 | 55 | 0 | 0 | 3 | 174 | 0 | 0.55 | 0.56 | 0.57 | 0.56 | 0.55 | 0.50 | 0.55 | −6.09 | −0.28 | −0.47 |
| rueresco | H | 21 | 61 | 0 | 21 | 4 | 579 | 0 | −1.13 | −7.39 | −10.30 | −10.73 | −8.10 | −13.26 | −17.25 | −35.85 | −10.54 | −15.36 |
| taylorpark | I | 21 | 62 | 0 | 50 | 8 | 658 | 50 | 0.67 | 0.86 | 0.76 | 0.75 | 0.79 | 0.37 | 0.51 | −2.60 | −0.12 | −1.66 |
| morrowpoint | FC | 21 | NA | 3 | 66 | 4 | 9420 | 42 | −1.36 | −0.70 | −1.30 | −2.00 | −0.75 | −2.21 | −2.25 | −1.31 | −14.36 | −32.07 |
| elephantbuttedam | FC | 22 | NA | 15 | 81 | 23 | 77,826 | 7 | −58.01 | 0.75 | 0.72 | 0.48 | 0.75 | −2.21 | −16.85 | 0.74 | −86.27 | −132.15 |
| lvks | FC | 27 | 9 | 0 | 66 | 12 | 896 | 65 | 0.18 | 0.66 | 0.54 | −4.64 | 0.63 | −0.78 | −0.45 | 0.66 | −0.01 | −0.05 |
| hsne | FC | 27 | 9 | 0 | 0 | 7 | 2295 | 0 | −3.03 | −154.90 | −160.69 | −166.04 | −153.99 | −154.89 | −418.46 | −147.06 | −105.60 | −136.25 |
| gibr | I | 41 | 55 | 0 | 92 | 5 | 1436 | 0 | 0.76 | 0.93 | 0.47 | 0.39 | 0.82 | 0.48 | 0.54 | 0.62 | 0.26 | 0.16 |
| sher | I | 41 | 68 | 0 | 47 | 5 | 171 | 23 | 0.74 | 0.95 | 0.65 | 0.53 | 0.91 | 0.53 | 0.52 | 0.75 | 0.42 | 0.30 |
| pist | FC | 42 | 17 | 0 | 68 | 8 | 2342 | 0 | −0.08 | −0.22 | 0.09 | 0.10 | 0.07 | −0.64 | 0.16 | 0.06 | −13.14 | −29.56 |
| ler | FC | 42 | 24 | 10 | 62 | 60 | 11,378 | 0 | −1.48 | −1.00 | −0.66 | 0.52 | −1.01 | 0.59 | 0.49 | −1.05 | 0.57 | 0.51 |
| shr | FC | 43 | 14 | 1 | 83 | 21 | 7751 | 0 | 0.21 | −1.96 | −0.23 | −0.13 | −2.81 | −0.75 | −0.56 | −1.59 | −15.27 | −224.52 |
| ltr | FC | 43 | 17 | 1 | 89 | 13 | 4438 | 9 | −0.20 | −0.15 | 0.01 | −0.14 | 0.01 | −0.45 | −0.06 | 0.00 | −2.07 | −35.12 |
| hbne | FC | 44 | 10 | 0 | 0 | 6 | 1775 | 0 | −26.11 | −159.76 | −163.08 | −160.60 | −157.51 | −217.66 | −389.11 | −155.39 | −304.84 | −371.83 |
| jamr | FC | 46 | 18 | 6 | 65 | 18 | 4132 | 66 | −0.02 | −1.35 | −1.51 | −1.39 | −1.50 | −18.16 | −19.17 | −2.18 | −195.35 | −634.67 |

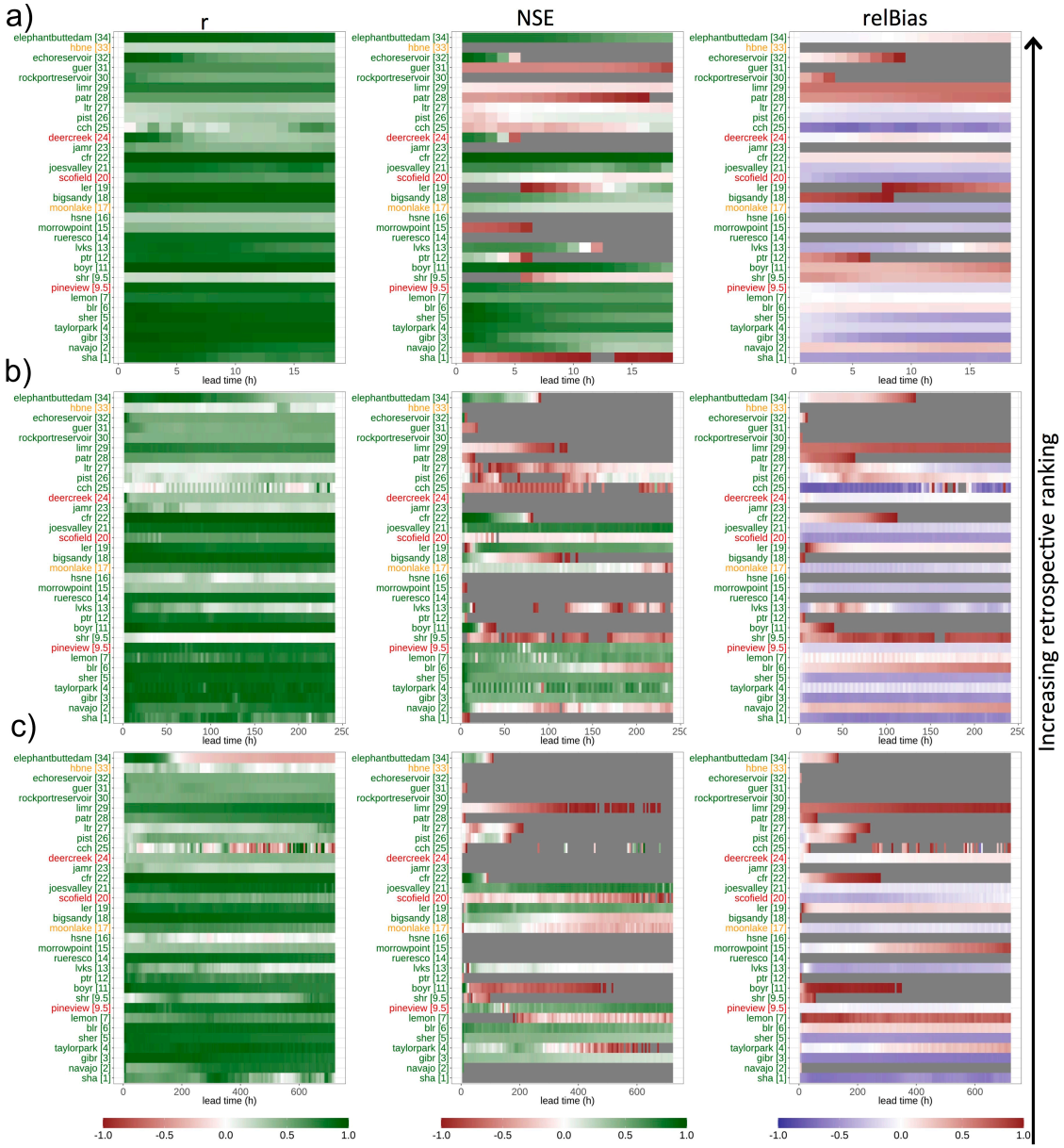



**Figure 7.** Performance of NWM reservoir inflows across lead times for one year, where reservoirs in each subplot are ordered from bottom to top by the retrospective ranking (best to worst, listed in parenthesis on the *y*-axis with reservoir name). Each column from left to right columns illustrate *r*, *NSE*, *relBias*, where the metric is scaled by color density and grey is out of valid range. The subplot rows are forecast cycles: short (**a**), medium (**b**), long (**c**).

Generally, NWM reservoir inflow performance between the forecast and retrospective agree, with some exceptions (Figure 7). One example of these exceptions is Navajo reservoir (*navajo*), which ranks high in the retrospective evaluation but has poor goodness of fit metrics across forecast cycles. This contrast between forecast and retrospective suggests issues with the meteorological forcings used to drive the forecasts. In Navajo, only 2% of the contributing watershed is calibrated, and 76% of the basin is gauged (high assimilation). The NWM simulates flows with high fidelity in the retrospective (Figure 8, panel a), but some loss in performance in the forecast for some possible forcing overestimation. The short range hydrograph is illustrated in Figure 8 panel b, while medium and long range are available in the additional material in Figure A4. The converse is true for Elephant Butte (*elephantfbutte*), which has a low retrospective performance (see ranking in Figure 7 and hydrograph in

Figure 9, panel a), but a relatively high goodness of fit metrics in the forecast short lead times (Table 3 and hydrograph at Figure 9, panel b (short range, while medium and long range are available in additional material at Figure A5). In this case, we note that *elephantbutte* has 15 lakes upstream and significant diversions in the basin; and 81% of the basin is gauged, which means assimilation is a key factor in achieving an $NSE = 0.75$ in early lead times across all forecast cycles in this basin (Table 4). For longer lead times, conversely, the effects of assimilation vanish, and we assist to a rapid decrease in performance at longer lead times (Table 4 and medium- and long-range hydrographs for *elephantbutte* in AX3). Table 4 indicates that more often, simulated reservoir inflows perform worse in the retrospective and better in the forecast than the opposite case. These reservoirs also have a considerable amount of assimilation upstream, proving the effectiveness of nudging to improve performance, at least for some of the cycles.

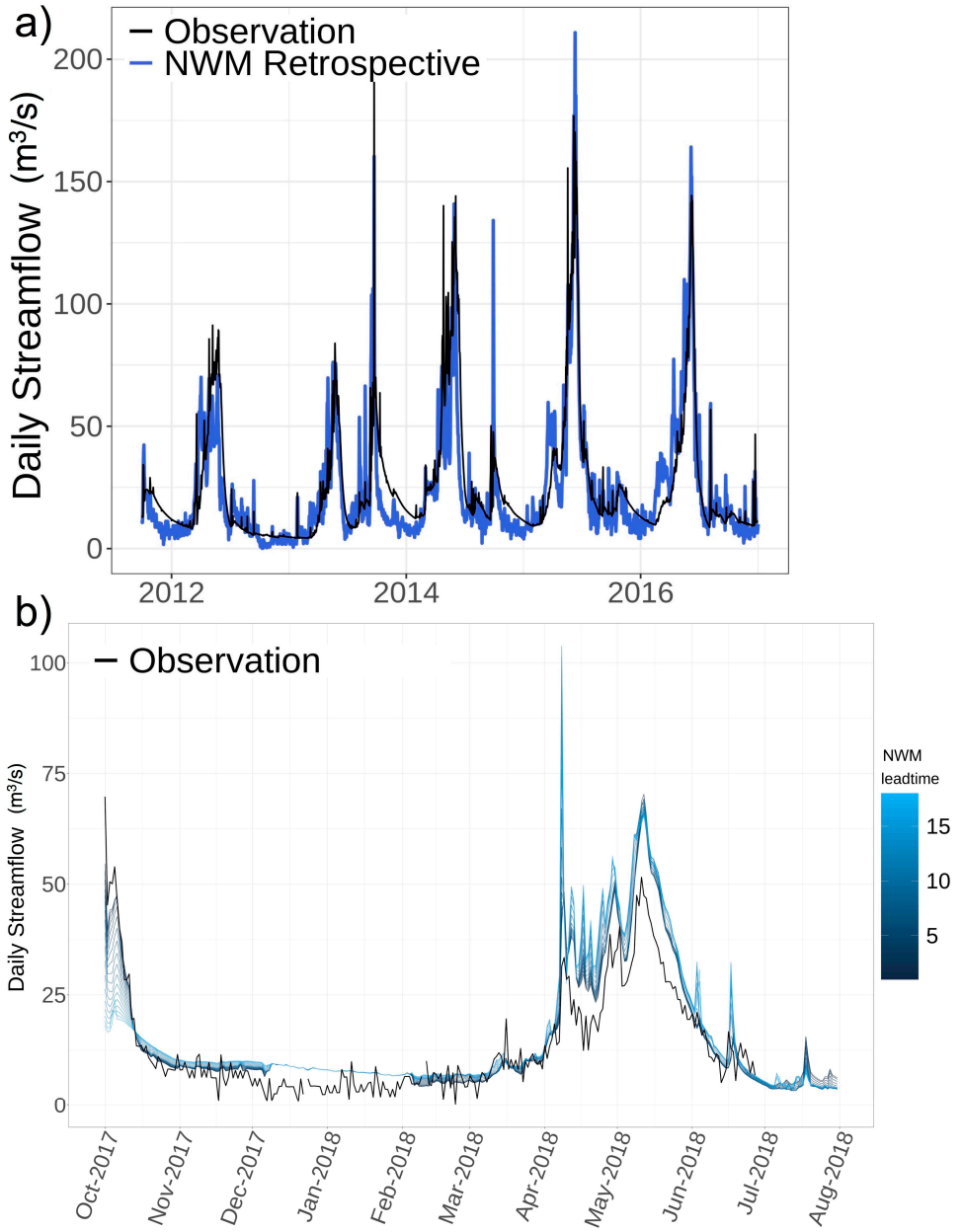



**Figure 8.** NWM simulated reservoir inflows into Navajo Dam: (**a**) shows the NWM retrospective simulation (v1.2 in blue) against observations (black) for a 5-year period, (**b**) shows the Navajo Dam simulated reservoir inflows for a one-year forecast period, where the blue lines are the modeled flows for all lead times and black is the BoR observations.

This analysis suggests that certain reservoir inflows are reliable for the retrospective and forecast periods, such as *lemon*, *blr*, *sher*, *taylorpark*, *gbr*, while others may be more suitable to provide information only for a particular forecast cycle.

Table 4 also highlights that NWM performs relatively well in snow-driven watersheds, in both retrospective and forecast simulations. In this analysis, the snow-driven watersheds tend to be small, headwaters basins, with minimal upstream management, and well calibrated with a high percentage of gauged (assimilated) area (e.g., *taylorpark*, *blr*, *sher*). In such basins, where the hydrology is tied to seasonal accumulation and snow melt periods rather than convective storms, the NWM forecasts and retrospectives can be useful for planning purposes such as those described in Table 2.

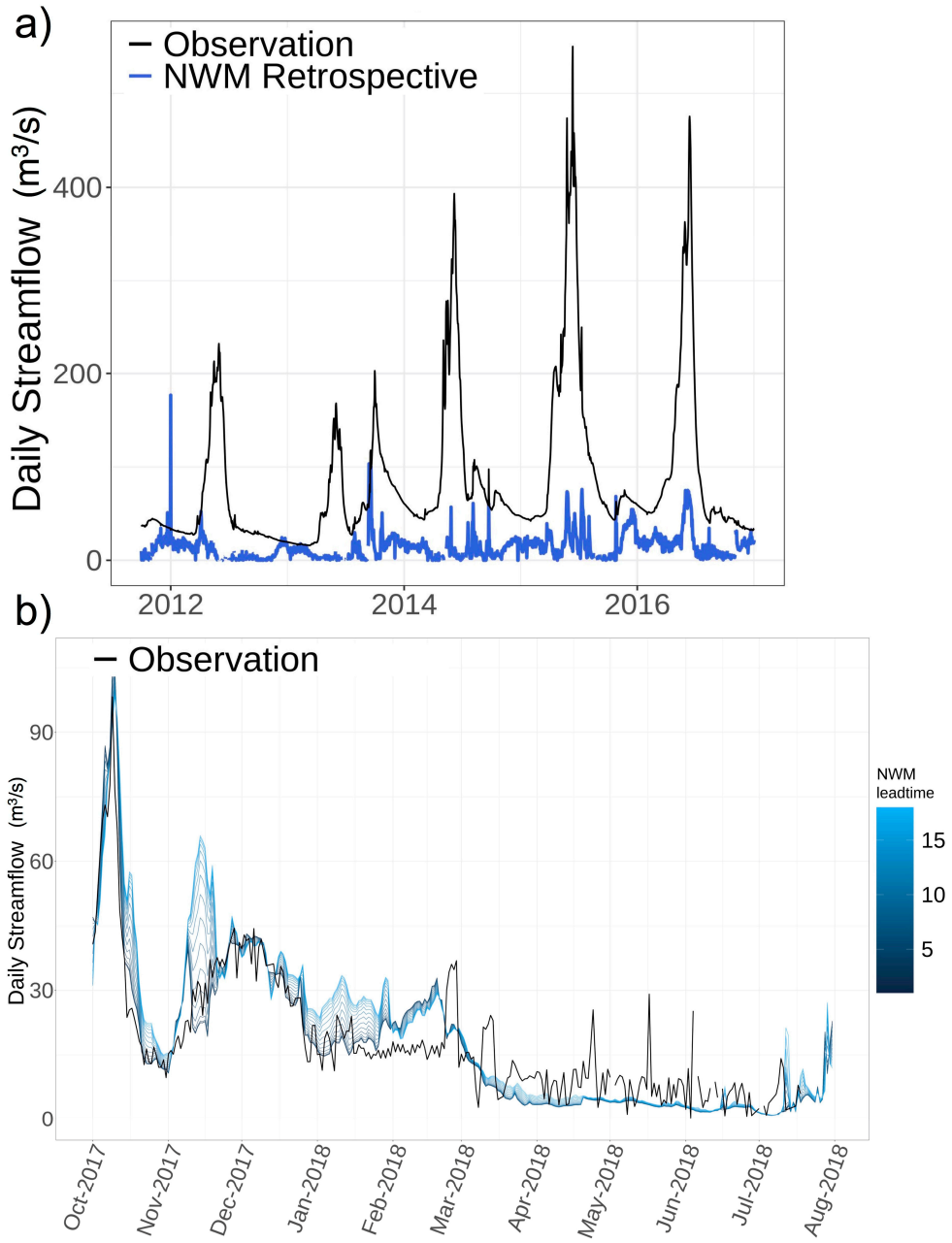



**Figure 9.** NWM simulated reservoir inflows into Elephant Butte Reservoir: (**a**) illustrates the NWM retrospective simulation (v1.2 in blue) against observations (black) for a 5-year period; (**b**) shows the Elephant Butte Reservoir simulated reservoir inflows for a one-year forecast period, where the blue lines are the modeled flows for all lead times and black is the BoR observations.

## 4. Discussion

In general, we found that simulated inflows into reservoirs with upstream management have more bias and lower *NSE* and r skill scores in the retrospective and forecast evaluation compared to those in unregulated watersheds. Furthermore, as the number of upstream lakes increase, performance in the forecast decays rapidly as a function of lead time. As explained in Section 2, the NWM currently employs a simplified representation of management from upstream reservoirs and this is, as expected, reflected in poorer performances both in the retrospective and in the forecast cycles for watersheds that include a high number of regulated reservoirs and lakes upstream.

The performance of reservoirs by watershed size is highly correlated with the number of lakes: most of the time a bigger water projects include multiple upstream projects/facilities and, consequently, the NWM's simulation of natural processes cannot capture these features.

As with the retrospective evaluation, we find that basins with partially calibrated drainage areas generally have worse goodness of fit scores than fully calibrated across all forecast cycles, an indication that regionalization within a large watershed may adversely impact performance. For example, if a large watershed has multiple regionalization "clusters" within it, the parameters may be distributed in a non-consistent manner across the different contributing basins to the total inflow.

In addition to that, the joined forecast and retrospective evaluation of reservoirs in Table 4 highlights some important considerations regarding the role of assimilation and reinforces the discussion of Figure 4: some reservoirs that have very poor performance in the retrospective and large assimilated area are able to have much better performance in the short lead times (e.g., Pactola reservoir (*ptr*), Echo reservoir (*echoreservoir*), Elephant Butte Dam (*elephantbuttedam*), Lovewell Dam (*lvks*)) presumably thanks to correction provided by assimilation.

Finally, another important surprising result is the watersheds with a higher percentage of snow precipitation performed better than the ones fed mostly by rain on both retrospective and forecast evaluation. Although it is beyond the scope of this study to conduct specific experiments related to snow-driven processes, it was still possible to state the fact that the current NWM has been better optimized to reproduce the runoff characteristics in these snow-driven watersheds. Even if this can also sum to the ability of the forcings to correctly depict solid vs. liquid precipitation, these results are very consistent between retrospective and forecast, suggesting that rain-fed watersheds are more challenging for the NWM to simulate, probably due to flashy behavior in the Midwestern U.S. plains compared to many snow-driven watersheds in the western U.S. Further research into better meteorological forcings and physics improvements are needed to address this issue.

## 5. Conclusions

In the absence of an observed total inflow into critical reservoirs, the NWM can provide an understanding of the hydrological processes and patterns of the contributing watershed hydrology from the retrospective simulations. The NWM also provides access to a CONUS-wide set of continuous forecasts which use consistent model physics and meteorological forcings at near real-time, which can serve as an additional set of information for water managers. As NWM evolves from its nascent stages, there is a great need and effort to encourage participation in model improvements and evaluating sources for data assimilation and parameter estimation. However, the ability to directly use the forecasted inflows depends on several factors. The main factors driving inflow performance are the precipitation dominance (snow vs. rain fed) and the influence of upstream management (number of lakes upstream). Other physical characteristics were influential but not in a consistent and systematic way. Thus, in headwater basins that are snow-driven, the NWM can provide a realistic depiction of the flow profile.

This analysis also exposed short-comings with regard to the mixed performance in calibration basins, which were expected to show improvements especially, in the retrospective. Given the variability in observation quality as denoted in Section 2.2, it is difficult to attribute poor performance strictly to the model; thus, future work will expand the number of reservoirs to include those from other

agencies that have inflow observations. Overall, even if local engineers will still use their methods for accounting total inflows in reservoirs of interest, in ungauged headwater basins, NWM provides information that can supplement these calculations. Moreover, the hourly time resolution offered by the NWM may be useful to operators and forecasters who otherwise might only have access to daily information. With a growing community of researchers and local offices that are getting involved in exploring the potential usage of NWM for water management, the hope is that this study will offer good guidance for future work and applications.

**Author Contributions:** Conceptualization, L.R. and F.V.; methodology, L.R. and F.V.; validation, F.V. and L.R.; formal analysis, F.V. and L.R.; resources, K.N., M.H., R.C.; writing—original draft preparation, L.R. and F.V.; writing—review and editing, F.V., L.R., R.C., M.H. and K.N.; supervision, D.G., A.W.W., M.H., R.C., K.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was made possible through support from the Bureau of Reclamation (BoR) Science and Technology Program (Project 1774).

**Acknowledgments:** We are grateful to the many BoR reservoir operators and NOAA River Forecast Center and NOAA PSL personnel who provided helpful information on reservoir inflows to augment the data provided through the BoR RISE database and contributed in matching reservoir purposes with NWM forecast ranges. In particular we want to thank for this Lynn Johnson (Cooperative Institute for Research in the Atmosphere (CIRA)/NOAA−Physical Science Laboratory (PSL)), Noe Santos (Bureau of Reclamation, Lower Colorado Basin Region), Peter Fickenscher (River Forecast Center, California Nevada River Forecast Center (CNRFC)), Jordan Lanini (Bureau of Reclamation).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| ARMA/ARIMA | Auto-regressive (Integrated) Moving Average |
| AWS | Access2Water |
| *ber* | Lake Berriessa and Monticello Dam |
| BoR | Bureau of Reclamation |
| CFS | Climate Forecast System |
| CONUS | Continental United States |
| CWMS | Corps Water Management System |
| *echoreservoir* | Echo Reservoir |
| *elephantbuttedam* | Elephant Butte Dam |
| GFS | Global Forecast System |
| HRRR | High−Resolution Rapid Refresh |
| *lvks* | Lovewell Dam |
| *navajo* | Navajo Reservoir |
| NCEP | National Centers for Environmental Prediction |
| NHD | National Hydrography Dataset |
| NID | National Inventory of Dams |
| NLDAS−2 | National Land Data Assimilation System |
| NOAA | National Oceanic Atmospheric Administration |
| *NSE* | Nash–Sutcliffe Efficiency |
| NWM | National Water Model |
| *ptr* | Pactola Reservoir |
| *r* | Pearson's correlation coefficient |
| *relBias* | Relative Bias |
| RFC | River Forecast Center |
| RISE | Reclamation Information Sharing Environment |
| *sher* | Lake Sherburne |
| USGS | United States Geological Survey |
| WFO | Weather Forecast Office |
| WRF | Weather Research Forecast |

## Appendix A

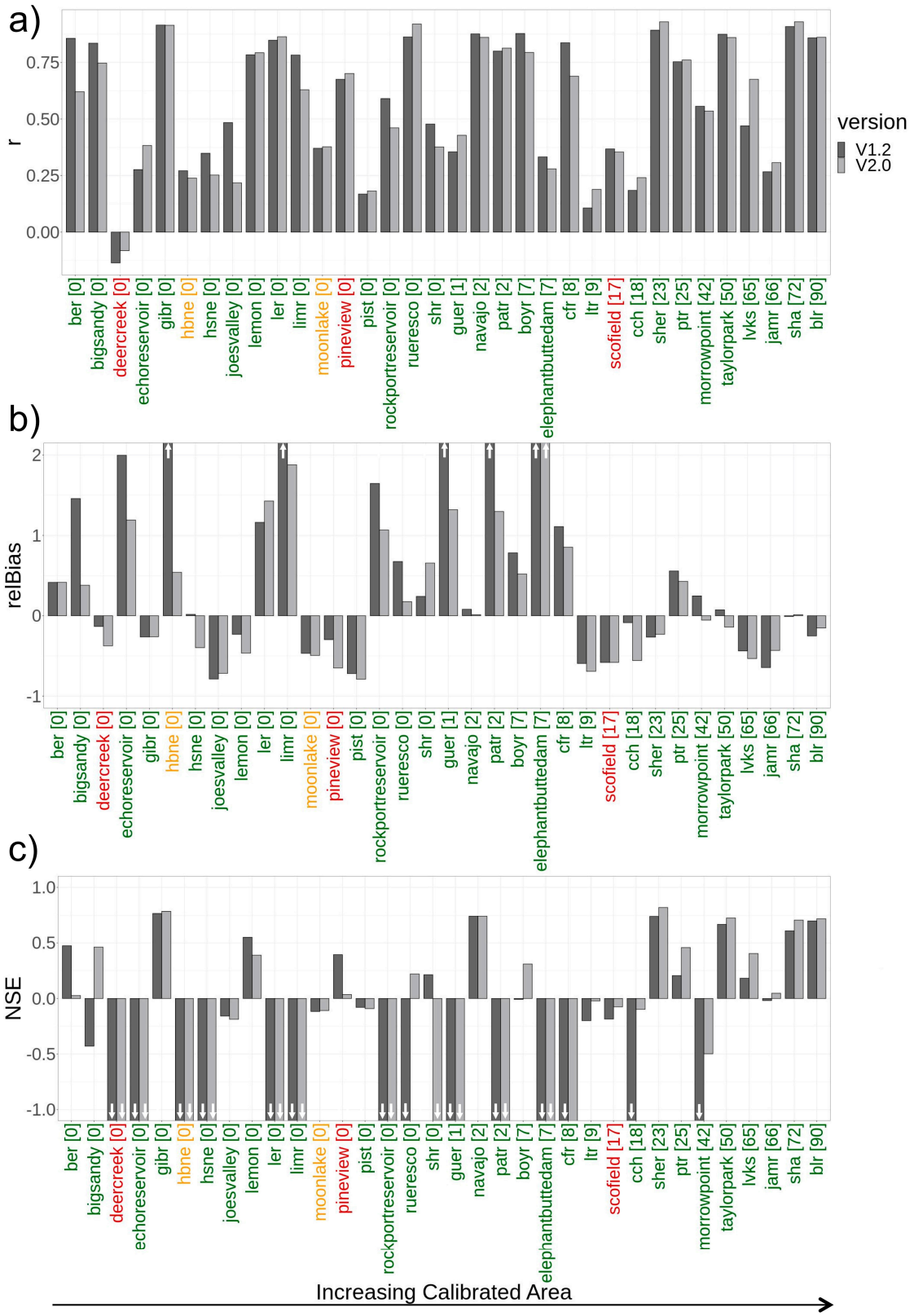



**Figure A1.** Goodness of fit scores from the NWM v1.2 (dark grey) and NWM v2.0 (light grey) retrospective total inflow evaluation, ordered by increasing calibrated area. From top to bottom: *r* (**a**), *relBias* (**b**), *NSE* (**c**).

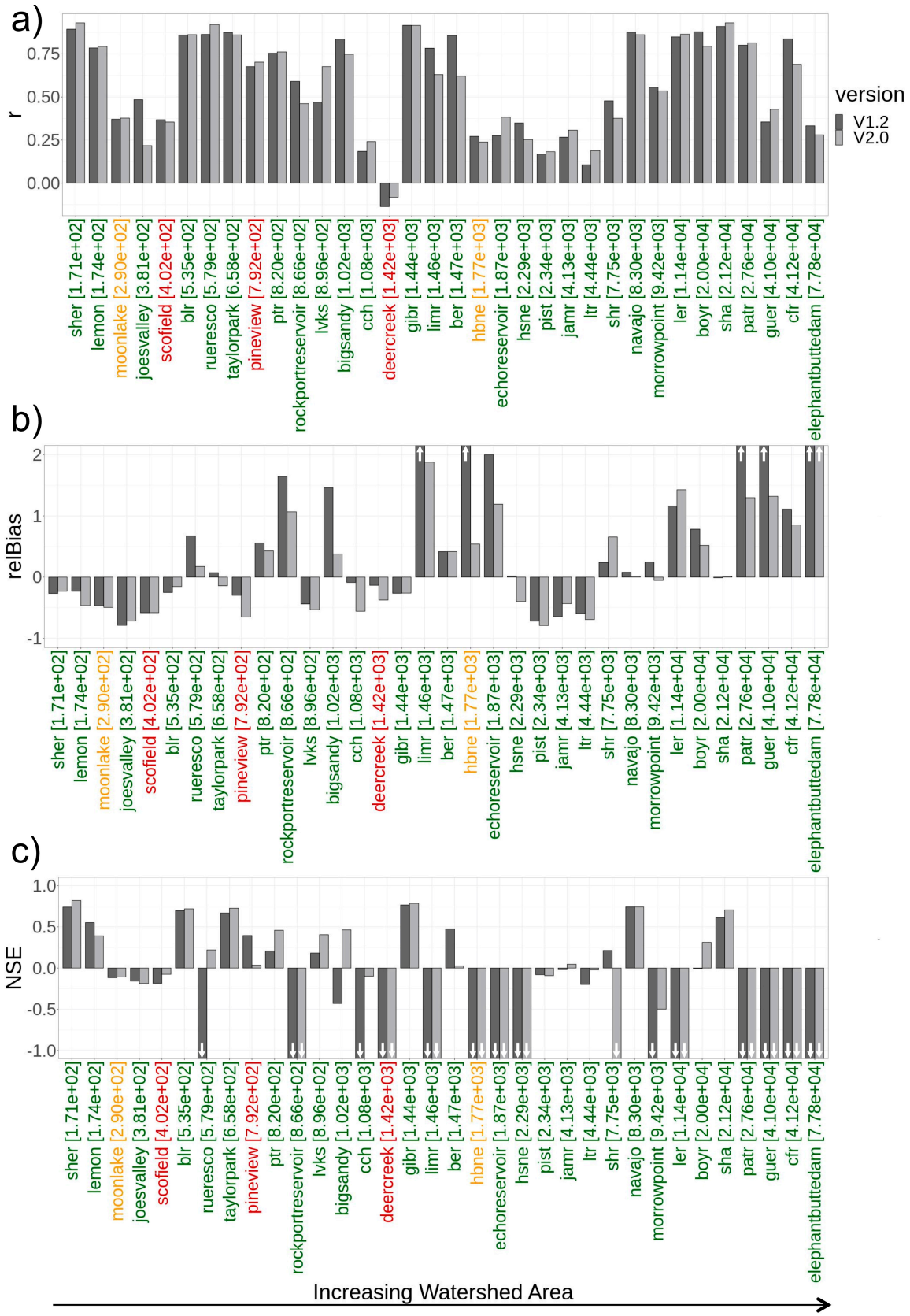



**Figure A2.** Goodness of fit scores from the NWM v1.2 (dark grey) and NWM v2.0 (light grey) retrospective total inflow evaluation, ordered by increasing watershed area. From top to bottom: *r* (**a**), *relBias* (**b**), *NSE* (**c**).

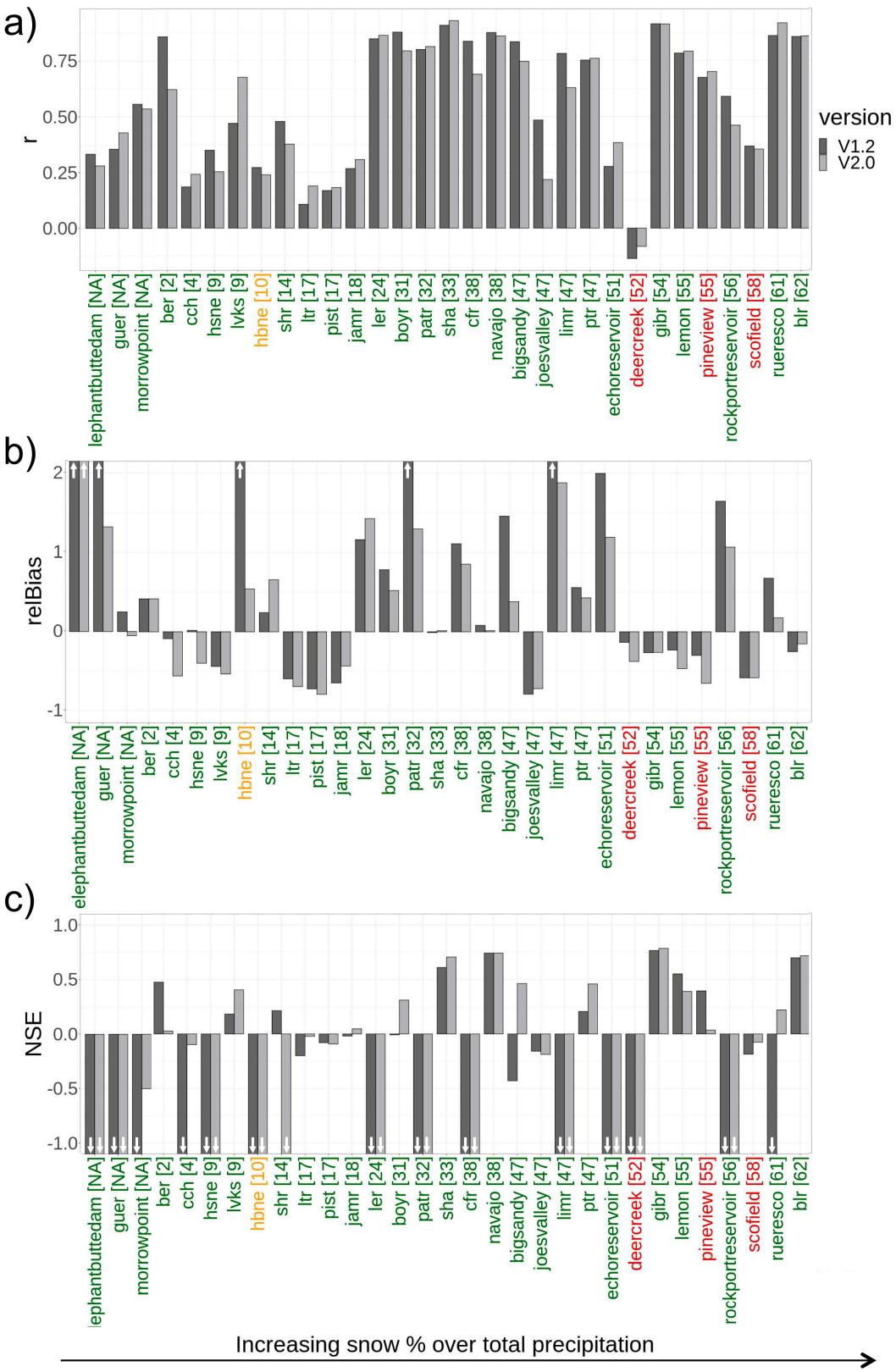



**Figure A3.** Goodness of fit scores from the NWM v1.2 (dark grey) and NWM v2.0 (light grey) retrospective total inflow evaluation, ordered by increasing snow percentage over total precipitation.from top to bottom: *r* (**a**), *relBias* (**b**), *NSE* (**c**). When the snow percentages are replaced by NAs means that that particular reservoirs did not have any GAGES2 snow percentage information.

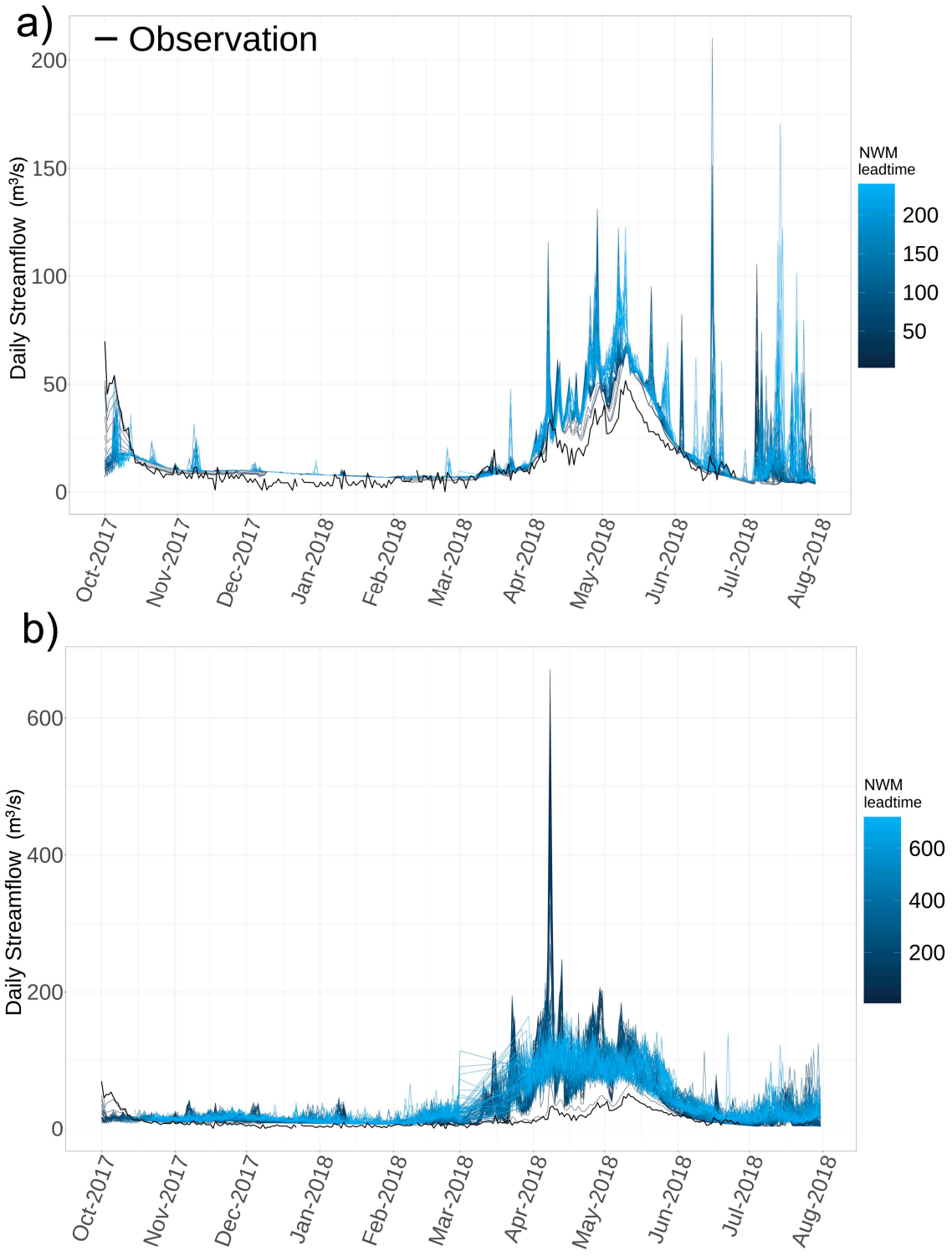



**Figure A4.** NWM simulated reservoir inflows into Navajo Dam for a one−year forecast period for medium (**a**) and long range (**b**). Blue lines are the modeled flows for all lead times and black is the BoR observations.

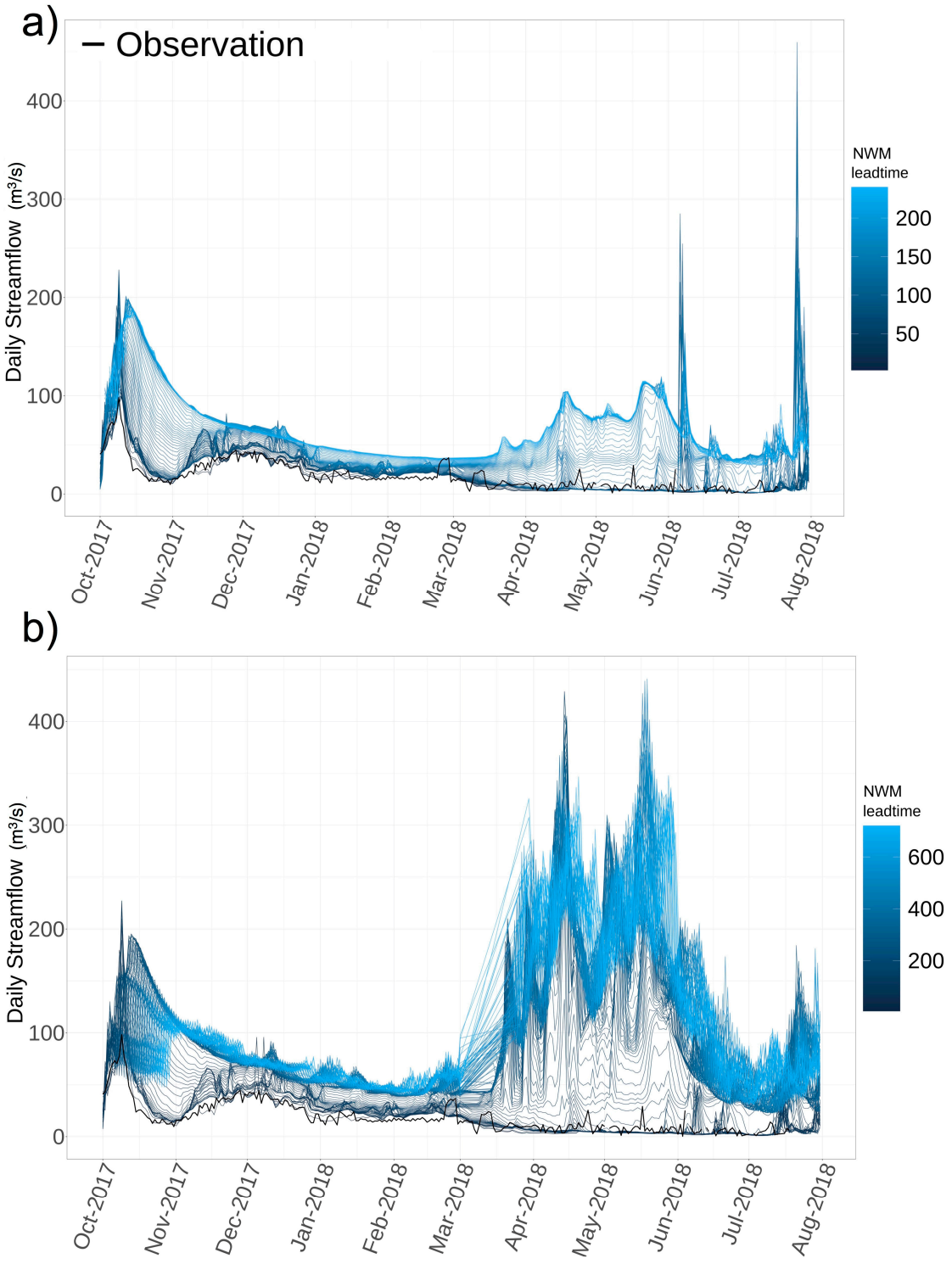



**Figure A5.** NWM simulated reservoir inflows into Elephant Butte Reservoir for a one−year forecast period for medium (**a**) and long range (**b**). Blue lines are the modeled flows for all lead times and black is the BoR observations.

## References

1. Turner, S.W.D.; Bennett, J.C.; Robertson, D.E.; Galelli, S. Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 4841–4859. [CrossRef]
2. Georgakakos, K.P.; Graham, N.E. Potential benefits of seasonal inflow prediction uncertainty for reservoir release decisions. *J. Appl. Meteorol. Clim.* **2008**, *47*, 1297–1321. [CrossRef]

3. Zhao, T.; Yang, D.; Cai, X.; Zhao, J.; Wang, H. Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast. *Water Resour. Res.* **2012**, *48*, 2012. [CrossRef]
4. Pretto, P.B.; Chiew, F.H.S.; McMahon, T.A.; Vogel, R.M.; Stedinger, J.R. The (mis)behavior of behavior analysis storage estimates. *Water Resour. Res.* **1997**, *33*, 703–709. [CrossRef]
5. Rippl, W. The capacity of storage-reservoirs for water-supply (including plate). *Minutes Proc. Inst. Civ. Eng.* **1883**, *71*, 270–278.
6. Hrachowitz, M.; Savenije, H.H.G.; Blöschl, G.; McDonnell, J.J.; Sivapalan, M.; Pomeroy, J.W.; Arheimer, B.; Blume, T.; Clark, M.P.; Ehret, U.; et al. A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrol. Sci. J.* **2013**, *58*, 1198–1255. [CrossRef]
7. Sivapalan, M. Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrol. Process.* **2003**, *17*, 3163–3170. [CrossRef]
8. Sivapalan, M.; Takeuchi, K.; Franks, S.W.; Gupta, V.K.; Karambiri, H.; Lakshmi, V.; Liang, X.; McDonnell, J.J.; Mendiondo, E.M.; O'Connell, P.E.; et al. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **2003**, *48*, 857–880. [CrossRef]
9. Deng, C.; Liu, P.; Liu, Y.; Wu, Z.; Wang, D. Integrated hydrologic and reservoir routing model for real-time water level forecasts. *J. Hydrol. Eng.* **2015**, *20*. [CrossRef]
10. Deng, C.; Liu, P.; Guo, S.; Wang, H.; Wang, D. Estimation of nonfluctuating reservoir inflow from water level observations using methods based on flow continuity. *J. Hydrol.* **2019**, *529*, 1198–1210. [CrossRef]
11. Anderl, B.; Attmannspacher, W.; Schultz, G.A. Accuracy of reservoir inflow forecasts based on radar rainfall measurements. *Water Resour. Res.* **1976**, *12*, 217–223. [CrossRef]
12. Bonnema, M.; Sikder, S.; Miao, Y.; Chen, X.; Hossain, F.; Ara Pervin, I.; Mahbubur Rahman, S.M.; Lee, H. Understanding satellite-based monthly-to-seasonal reservoir outflow estimation as a function of hydrologic controls. *Water Resour. Res.* **2016**, *52*, 4095–4115. [CrossRef]
13. Ali, S.A.; Sridhar, V. Deriving the reservoir conditions for better water resource management using satellite-based earth observations in the lower Mekong River basin. *Remote Sens.* **2019**, *11*, 2872. [CrossRef]
14. Choi, G.; Shin, H.; Yoon, S. Research on dam inflow analysis based on radar rainfall data. *EGU Gen. Assem. 2020* **2020**, *2116*, 5194.
15. Sivapalan, M. The secret to "doing better hydrological science": Change the question! *Hydrol. Process.* **2009**, *23*, 1391–1396. [CrossRef]
16. Razavi, T.; Coulibaly, P. Improving streamflow estimation in ungauged basins using a multi-modelling approach. *Hydrol. Sci. J.* **2016**, *61*, 2668–2679. [CrossRef]
17. Gupta, H.V.; Wagener, T.; Liu, Y. Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrol. Process.* **2008**, *22*, 3802–3813. [CrossRef]
18. van Emmerik, T.; Mulder, G.; Eilander, D.; Piet, M.; Savenije, H. Predicting the ungauged basin: Model validation and realism assessment. *Front. Earth Sci.* **2015**, *3*, 1–11. [CrossRef]
19. Peng, C.; Buras, N. Practical estimation of inflows into multireservoir system. *J. Water Resour. Plan. Manag.* **2000**, *126*, 331–334. [CrossRef]
20. Parajka, J.; Viglione, A.; Rogger, M.; Salinas, J.L.; Sivapalan, M.; Blöschl, G. Comparative assessment of predictions in ungauged basins—Part 1: Runoff-hydrograph studies. *Hydrol. Earth Syst. Sci.* **2013**, *10*. [CrossRef]
21. Zhao, T.; Cai, X.; Yang, D. Effect of streamflow forecast uncertainty on real-time reservoir operation. *Adv. Water Resour.* **2011**, *34*, 495–504. [CrossRef]
22. Turner, S.W.D.; Xu, W.; Voisin, N. Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1275–1291. [CrossRef]
23. Yeh, W.W. Reservoir Management and Operations Models: A State-of-the-Art Review. *Water Resour. Res.* **1985**, *21*, 1797–1818. [CrossRef]
24. Jain, S.K.; Das, A.; Srivastava, D.K. Application of ANN for reservoir inflow prediction and operation. *J. Water Resour. Plan. Manag.* **1999**, *125*, 263–271. [CrossRef]
25. Coulibaly, P.; Anctil, F.; Bobée, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* **2000**, *230*, 244–257. [CrossRef]
26. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* **2013**, *476*, 433–441. [CrossRef]

27. Lima, L.M.M.; Popova, E.; Damien, P. Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. *Int. J. Forecast.* **2014**, *30*, 464–476. [CrossRef]
28. Stedinger, J.R.; Sule, B.F.; Loucks, D.P. Stochastic dynamic programming models for reservoir operation optimization. *Water Resour. Res.* **1984**, *20*, 1499–1505. [CrossRef]
29. Faber, B.A.; Stedinger, J.R. Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. *J. Hydrol.* **2001**, *249*, 113–133. [CrossRef]
30. Castelletti, A.; Galelli, S.; Restelli, M.; Soncini-Sessa, R. Tree-based reinforcement learning for optimal water reservoir operation. *Water Resour. Res.* **2010**, *46*. [CrossRef]
31. Denaro, S.; Anghileri, D.; Giuliani, M.; Castelletti, A. Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data. *Adv. Water Resour.* **2017**, *103*. [CrossRef]
32. Stern, C.V.; Normand, A.E. *Bureau of Reclamation: History, Authorities, and Issues for Congress*; Report No. R46303 2020-04; Congressional Research Service: Washington, DC, USA, 2020; p. 24.
33. U.S. Department of Commerce; U.S. Department of Interior. *Federal Action Plan for Improving Forecasts of Water Availability*; Report from October 2019; U.S. Department of Interior: Washington, DC, USA, 2019.
34. Viterbo, F.; Mahoney, K.; Read, L.; Salas, F.; Bates, B.; Elliott, J.; Cosgrove, B.; Dugger, A.; Gochis, D.; Cifelli, R. A multiscale, hydrometeorological forecast evaluation of national water model forecasts of the may 2018 Ellicott City, Maryland, Flood. *J. Hydrometeorol.* **2020**, *21*, 475–499. [CrossRef]
35. National Centers for Environmental Prediction (NCEP). National Water Model Products. Available online: https://www.nco.ncep.noaa.gov/pmb/products/nwm/ (accessed on 9 August 2020).
36. Office of Water Prediction The National Water Model. Available online: https://water.noaa.gov/about/nwm (accessed on 16 August 2020).
37. Gochis, D.J.; Barlage, M.; Cabell, R.; Casali, M.; Dugger, A.; Fitzgerald, K.; Mcallister, M.; Mccreight, J.; Rafieeinasab, A.; Read, L.; et al. The WRF-Hydro Modeling System Technical Description; Version 5.1; NCAR Techinical Notes. 2020. Available online: https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf (accessed on 17 August 2020).
38. McKay, L.; Bondelid, T.; Dewald, T.; Johnston, J.; Moore, R.; Rea, A. *NHDPlus Version 2: User Guide*; U.S. Environmental Protection Agency: Washington, DC, USA, 2012.
39. Bureau of Reclamation. Reclamation Information Sharing Environment (RISE). Data Portal. Available online: https://water.usbr.gov/docs/RISE.pdf (accessed on 12 October 2020).
40. Kim, J.; Read, L.; Johnson, L.E.; Gochis, D.; Cifelli, R.; Han, H. An experiment on reservoir representation schemes to improve hydrologic prediction: Coupling the national water model with the HEC-ResSim. *Hydrol. Sci. J.* **2020**. [CrossRef]
41. Bureau of Reclamation. Water Operations: Historic Data. Available online: https://www.usbr.gov/rsvrWater/HistoricalApp.html (accessed on 8 September 2020).
42. Lukas, J.; Payton, E. Colorado River Basin Climate and Hydrology: State of the Science. Western Water Assessment. 2020. Available online: https://wwa.colorado.edu/publications/reports/CRBreport/ (accessed on 30 August 2020).
43. Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **2001**, *51*, 933–938. [CrossRef]
44. Gochis, D.; Yates, D.; Sampson, K.; Dugger, A.; McCreight, J.; Barlage, M.; RafieeiNasab, A.; Karsten, L.; Read, L.; Zhang, Y.; et al. Overview of National Water Model Calibration General Strategy & Optimization. Available online: https://ral.ucar.edu/sites/default/files/public/9_RafieeiNasab_CalibOverview_CUAHSI_Fall019_0.pdf (accessed on 9 August 2020).
45. Falcone, J.A. GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow. US Geological Survey Digital Spatial Dataset. 2011. Available online: https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml (accessed on 11 July 2020).
46. McCabe, G.J.; Wolock, D.M. Long-term variability in Northern Hemisphere snow cover and associations with warmer winters. *Clim. Chang.* **2010**, *99*, 141–153. [CrossRef]
47. U.S. Army Corps of Engineers National Inventory of Dams. Available online: https://www.publications.usace.army.mil/Portals/76/Publications/EngineerPamphlets/EP_360-1-23.pdf (accessed on 16 July 2019).

48. Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]
49. Hazewinkel, M. (Ed.) *Encyclopaedia of Mathematics: C An Updated and Annotated Translation of the Soviet Mathematical Encyclopaedia*; Springer Science & Business Media: New York, NY, USA, 2013; Volume 2.
50. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
51. Van der Vaart, H.R. Some extensions of the idea of bias. *Ann. Math. Stat.* **1961**, *32*, 436–447. [CrossRef]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).