In mountainous areas, high-resolution atmospheric models can represent total annual precipitation better than the collective network of precipitation gauges.

We have now crossed a threshold where, for many mountain ranges, well-configured high-resolution atmospheric models are better able to represent range-wide total annual precipitation than the collective network of precipitation gauges: that is, observations. The prior sentence is disturbing. If we even assign some truth to the statement “Models are better than observations,” where does that lead us? If two models are “better than observations” but disagree with each other, which one should we trust more? What do we mean by “better,” and what counts as an “observation”? Generalities are dangerous, and for which models, which observations, and which times and locations might this statement be true? How do we identify these specificities in a way that allows us to move forward, scientifically and objectively, in a situation where the truth is difficult to discern?

Here, we review recent research that collectively suggests, at least for the mid- to northern latitudes, that modeled precipitation has crossed a threshold in range-wide accuracy relative to observation-based precipitation datasets in complex terrain. We carefully examine the basis for this conclusion, which often consists of multiple indirect observations. We then propose that crossing this threshold requires a fundamental shift in how hydrologists and atmospheric scientists interact. In the past, gridded precipitation datasets that interpolated between existing observations, such as PRISM (Daly et al.
provided a stable medium of both the best available input to a hydrologic model and the best available benchmark for an atmospheric model (Fig. 1). These gridded datasets are often referred to as observations, when in truth they are based on a statistical model interpolating between point measurements. To improve atmospheric model performance beyond the quality of a gridded precipitation field, it is necessary to explore other measurements, such as streamflow or snow on the ground. Similarly, for hydrologists to make better use of atmospheric model output, they need to understand how those models work, which requires greater knowledge of atmospheric science to understand what such models can and cannot be relied on for. Both steps require greater integration across disciplines to move forward.

**REVIEW OF TRADITIONAL DISCIPLINARY KNOWLEDGE OF OROGRAPHIC PRECIPITATION.** Why we care. Accurate precipitation estimates are essential for hydrologic predictions, ecological assessments, and infrastructure management. Hydrologists must translate atmospheric forecasts or point gauge measurements into precipitation amounts for the entire basin of interest. The importance of mountain precipitation to floods and summer water supplies has driven extensive studies and additional measurements in the fields of both hydrology and meteorology (Dettinger et al. 2004; Galewsky and Sobel 2005; Heggli and Rauber 1988; Jeton et al. 1996; Marwitz 1983; Pandey et al. 1999; Parish 1982; Ralph et al. 2005; Reeves et al. 2008; Smith et al. 2010). Despite our best efforts, many factors conspire to limit our ability to measure mountain

---

**Fig. 1.** Gridded precipitation products, such as PRISM (shown here for three river basins in California), have long served as the intermediary between hydrologic science and atmospheric science. As model output improves beyond the skill of our observational network and interpolation techniques, more complex and interdisciplinary evaluations are required.
precipitation, including blocked radar signals, spatial
details too fine for satellites or sparse gauge networks,
and precipitation gauges either capped by snow or
missing snow blown over the top of the orifice. Direct
measurements of precipitation, particularly snowfall,
in complex terrain are frequently scarce relative to the
spatial heterogeneity they are expected to measure and
may be unreliable (Derin et al. 2016; Milly and

Hydrologist’s approach to orographic precipitation.
Consistent and reliable estimates of precipitation are
essential to hydrologic models (Clark and Slater 2006;
Mizukami and Smith 2012). In situ data are considered
the gold standard, more accurate and precise than ei-
ther radar- or satellite-based estimates. The standard
practice in mountain hydrologic modeling is to dis-
tribute precipitation across a basin by some interpo-
lation scheme (Clark and Slater 2006; Daly et al. 1994;
Hay and Clark 2003). In flat terrain, radar signals aid
interpolation (Nelson et al. 2016), but in complex ter-
rain, radar signals are generally blocked, and with the
exception of a high-quality radar system in Switzerland
(e.g., Panziera et al. 2018), radars frequently report less
than 50% of observed precipitation (Ralph et al. 2014;
Trapero et al. 2009; Westrick et al. 1999; Young et al.
1999; Zhang et al. 2012). Interpolation can be done
either by allowing the number of stations and spatial
patterns to change through time (Clark and Slater
2006), or based on a climatological pattern assumed
to remain fixed proportional to a base station [e.g.,
Daly et al. (1994) and all datasets relying on PRISM,
see Table 1 in Lundquist et al. (2015)].

Both interpolation methods have problems—the
former due to hard-to-identify errors in high-eleva-
tion stations (e.g., Mizukami and Smith 2012) and
poorly characterized patterns on the event time scale,
and the latter due to spatial patterns of precipitation
that differ from climatology on both storm-specific
and even annual time scales (Lundquist et al. 2010,
2015). Efforts to avoid ubiquitous errors have used
runoff observations and evapotranspiration estima-
tions to “adjust” global precipitation datasets (Adam
et al. 2006; Fekete et al. 2002; Milly and Dunne 2002;
Xia 2008), with areas of limited gauge coverage and/
or severe gauge undercatch, such as the arctic, noted
as particularly problematic (Louie et al. 2002; Tian
et al. 2007; Ye et al. 2012). While gridded datasets
match precipitation observations well at locations
that have been incorporated into the interpolation
algorithms, most perform significantly less well at
locations in between stations, where no training data
were available (Currier et al. 2017; Gutmann et al.
2012; Hiemstra et al. 2006). Annual precipitation
from many gridded datasets does not even exceed
observed runoff in some mountain basins (Henn et al.
2018b); in their snow reanalysis approach Margulis
et al. (2015) must adjust incoming solid precipitation
to match high-elevation snow observations.

In addition to quantity, hydrologists often infer
precipitation phase from an empirical temperature
threshold guided by a gridded product of near-surface
temperature and sometimes humidity (Harpole et al.
2017; Jennings et al. 2018). The exact parameters in
these empirical algorithms are not transferable in
space and/or time (Jennings et al. 2018) and are of-
ten ill suited to changing synoptic conditions, since
conditions aloft also affect precipitation phase at
the surface (Wayand et al. 2016). Additionally, the
underlying temperature datasets may have signifi-
cant problems (Feld et al. 2013; Minder et al. 2010;

Atmospheric scientist’s approach to orographic precipi-
tation. While most hydrologists view precipitation as
an input parameter with some degree of uncertainty,
 atmospheric scientists view precipitation as the
output of dynamic and physical processes (Fig. 2).
Precipitation-generation mechanisms over terrain
include the lifting of moist air as a result of synoptic-
scale patterns (Fig. 2a), thermally driven convection,
and orographic lifting when air impinges on topogra-
phy (Fig. 2b), with multiple review summaries (Barros
and Lettenmaier 1994; Houze 2012; Roe 2005; Smith
1979). Orographic precipitation is modified by the
moisture flux, the slope of the terrain (Alpert 1986;
Neiman et al. 2009, 2013; Roe and Baker 2006; Smith
1979), and degree of blocking (Hughes et al. 2009;
Lundquist et al. 2010), as well as cloud microphysical
processes (Fig. 2c) (Grubišić et al. 2005; Jankov et al.
2009; Roe and Baker 2006; Yang et al. 2012), which
in turn affect the transition between hydrometeor
types: snow, ice/graupel, and rain (Fig. 2d); (Minder
and Kingsmill 2013; Minder et al. 2011). Upper-level
wind speeds and directions control where precipita-
tion reaches the surface, and complex wind patterns
near the mountain surface determine the patterns of
final snow distribution on the landscape (Greene et al.
1999; Mott and Lehning 2010; Winstal et al. 2002).

All of the above processes, with the exception of
finescale snow deposition and redistribution (Liston
et al. 1998; Vionnet et al. 2018), are incorporated into
modern numerical weather prediction models, which,
when using a grid spacing of 12 km or less, are able to
resolve the topography that drives major orographic
precipitation gradients (Anders et al. 2007; Barros

and Lettenmaier 1994; Liu et al. 2011). These models provide information for both short-term weather forecasts (Benjamin et al. 2016a; Mass et al. 2002), and downscaled climate projections (Liu et al. 2017; Rasmussen et al. 2019; Salathé et al. 2008).

Mesoscale model precipitation estimates can vary substantially depending on model configuration: for example, boundary layer, convection, and microphysics schemes (Hughes et al. 2019; Jankov et al. 2009; Morales et al. 2018; Yang et al. 2012) and boundary conditions (Hughes et al. 2019). Verification and configuration of these models relies heavily on observations; however, point observations may not be comparable to the model grid scale, and small space–time offsets in the model can appear as large errors (Casati et al. 2008; Cassola et al. 2015), a problem that grows more pronounced with higher-resolution models (Cassola et al. 2015; Lack et al. 2010). Gridded precipitation products are sometimes used to mitigate these mismatches (Ikeda et al. 2010; Liu et al. 2011; Xu et al. 2018), but errors in the statistically interpolated grid can also make results misleading (Henn et al. 2018b; Prein and Gobiet 2017).

MODEL VERSUS OBSERVATIONAL PRECIPITATION SKILL. **Gauge-based assessments.** Rain gauges undercatch actual rainfall in most environments (Collados-Lara et al. 2018; Liljedahl et al. 2017; Rodda 1968; Rodda and Dixon 2012; Sieck et al. 2007), and snowfall is even harder to measure (Rasmussen et al. 2012). Gridded precipitation products consistently match gauge-based precipitation better than atmospheric models, but given that the gridded products are interpolations between these very observations, this is not an independent comparison. At independent observation sites, models have generally outperformed gridded estimates for annual precipitation totals by a factor of 2 (Currier et al. 2017; Gutmann et al. 2012), although very few studies have compared with observational stations not used in statistical training, and the best performer may vary between years (Wayand et al. 2013). Compared to snow accumulation measured at snow pillows across California, gridded datasets were unbiased on average, but underpredicted (by as much as 50%) events with a large proportion of postfrontal precipitation (Lundquist et al. 2015). Henn et al. (2018a) demonstrated large spread (typically ±20% in annual means) between gridded datasets across the western United States, and multiple European studies found spread across gridded datasets was as large as, if not larger than, spread between regional climate models, with the greatest spread in areas of low gauge density (Herold et al. 2017; Isotta et al. 2015; Prein and Gobiet 2017). Zhang and Anagnostou (2019) found that mesoscale model output worked better than gauge-based precipitation observations for bias-correcting and downscaling satellite-based precipitation products during convective events in mountains in Colombia, Peru, and Taiwan.

Assessments using a hydrological or land surface model. Multiple studies have demonstrated that mesoscale model input may be comparable or preferable compared to gauge observations to drive a hydrologic and/or snow model in complex terrain. The majority of studies concluded that hydrologic model performance was similar between the two forcing datasets,
including areas in northwest Montana (Leung et al. 1996), the Pacific Northwest (Currier et al. 2017; Wayand et al. 2016), Colorado (Rasmussen et al. 2011), Northern California (Anderson et al. 2002), and Japan (Yoshitani et al. 2009). In the Pacific Northwest (Wayand et al. 2013; Westrick et al. 2002; Westrick and Mass 2001) and Iceland (Rodnvaldsson et al. 2004, 2007), mesoscale model forced simulations outperformed those using gauge data, which the authors attributed to unrepresentative gauge locations and gauge undercatch, respectively. A few studies found that gauge observations produced better streamflow results, but authors attributed this result to forecast errors because the mesoscale model was run in forecast mode (Anderson et al. 2002; Westrick et al. 2002) or to calibration bias because the hydrological model was calibrated to perform best when using gauge data (Kunstmann and Stadler 2005). A few studies demonstrated skillful mesoscale-model-forced snow simulations compared to snow observations, but did not directly compare with another simulation forced using gridded observations, including studies focused on Colorado (Ikeda et al. 2010; Rasmussen et al. 2011), the California Sierra Nevada (Wrzesien et al. 2015, 2017), and all of North America (Wrzesien et al. 2018). While all of the above results were also sensitive to the configuration and parameters of the selected models, they demonstrate that mesoscale model output is a viable option for mountain precipitation input in hydrologic applications.

To overcome the model dependence of the studies discussed above, some work has focused on a more direct way to extract precipitation from streamflow, while explicitly representing parameter uncertainties. The idea of doing “hydrology backwards” (Kirchner 2009) to infer precipitation from streamflow records has been formalized in a Bayesian framework (Kavetski et al. 2003; Kavetski et al. 2006a,b) and has been applied in a number of settings (Koskela et al. 2012; Kuczera et al. 2006; Renard et al. 2010; Thyer et al. 2009; Vrugt et al. 2008). These efforts attempt to account for uncertainties in other hydrologic fluxes, for example, evapotranspiration and groundwater gains and losses. In these examples, precipitation input uncertainty was accounted for by a set of precipitation multipliers used to relate gauge measured precipitation with basin average precipitation. These multipliers were inferred in conjunction with internal hydrologic model parameters, using Markov chain Monte Carlo (MCMC) iterations. Multiple studies (Henn et al. 2015, 2016, 2018b,c) employed these techniques and demonstrated that precipitation inferred from streamflow and snow observations shows greater spatial and temporal variability than gridded datasets, including those gridded products that aim to explicitly represent uncertainty (Newman et al. 2015), and that modeled precipitation better matches these estimates than a range of gridded datasets in the California Sierra Nevada (Hughes et al. 2019). Collectively, this work demonstrated that precipitation uncertainty was the dominant source of model error (Kavetski et al. 2006a; Kuczera et al. 2006).

A PATH FORWARD: HOW TO FURTHER IMPROVE. One of the most alarming conclusions from the cited works is that we cannot trust data-sets that are often considered truth. In most cases, problems arise from treating a statistical gridding of limited observations as actual observations. While our direct observations are sometimes flawed, they are the closest we can get to truth. Thus, our strategy should not be to abandon observations in favor of modeling, but rather, to focus on 1) obtaining the best observations possible, including investing in quality controlling those observations so that erroneous measurements are excluded from networks, 2) including all types of related observations (Fig. 3), and 3) working across disciplines to clearly communicate what each observation represents well, its uncertainty, and under which conditions it struggles. These principals are not new; however, the reality of working out what constitutes an error remains an enormous challenge (Dee 2005; Diamond et al. 2013).

**Quality measurements in the mountains.** We must maintain funding for our core observational networks, and provide increased support for their maintenance and, where and when possible, their expansion. In citizen science efforts, such as CoCoRaHS (Cifelli et al. 2005) or Community Snow Obs ([http://communitysnowobs.org/](http://communitysnowobs.org/)), more effort could be made to target people living in remote areas where gauge density is low, or to target areas beyond traditional roads, for example, ski huts, wilderness shelters, or fire lookouts.

However, incorrect observations are worse than no observations, and any effort to expand gauge networks must include adequate maintenance and quality control. River forecast centers regularly check for poor observations, but more effort could be invested into ways to automate such flagging of suspect observations due to their inconsistency with a model (e.g., see Hughes et al. 2012, their Fig. 9) or a range of model outputs, for example, testing of “hydrological coherence” (Laiti et al. 2018).
Ancillary measurements: Remote sensing, strategic radar, ecology, vapor flux, soil moisture, total storage. Significantly expanding direct observations of precipitation into remote complex terrain is likely not practical or feasible in the short term for many areas. However, models can be evaluated and improved with multiple other measurements that may be more robustly obtained.

Unlike most ground-based radar (Young et al. 1999), satellite views are not blocked by the mountains, and satellite flight frequency and resolution have been increasing in recent years (Entekhabi et al. 1999). The Tropical Rainfall Measurement Mission (TRMM) and the Global Precipitation Measurement (GPM) mission have provided estimates of precipitation over mountains where few in situ measurements exist, providing estimates of liquid precipitation (Houze et al. 2015), characteristics of convective precipitation (Rasmussen et al. 2013), and some options for hydrologic modeling (Xue et al. 2013). However, care must also be taken in the interpretation of satellite data, because these are modeled based on radiance retrieved at multiple wavelengths. To date, satellite algorithms have limited capability for atmospheric river events (Wen et al. 2018) or for estimating mixed or solid precipitation (Cao et al. 2018; Yong et al. 2012), particularly over complex terrain (Ebtehaj and Kummerow 2017; Kummerow et al. 2015). Globally, products that merge gauge data and satellite-based precipitation are often biased by unrepresentative and limited gauge data, particularly in mountains (Derin et al. 2016).

Progress has been made on better observing storm characteristics and vapor fluxes along mountain boundaries. For example, California has added upstream vertically pointing radars of various frequencies, including wind profilers and snow level radars (Ralph et al. 2014). Aircraft-based profile observations are also increasingly available, for example, aircraft meteorological data reports (AMDaR) and Tropospheric Aircraft Meteorological Data Reports (TAMDAR), (Moninger et al. 2003, 2010). These data sources could be assimilated to improve model predictions downwind or could be used to objectively select better performers from a suite of model options.

Although datasets that combine radar with data from other sources, such as the Multi-Radar Multi-Sensor (MRMS) product (Zhang et al. 2016), suffer from the same biases as their underlying datasets in complex terrain (Bytheway et al. 2019), gap-filling radar with a narrow beamwidth (e.g., C band or X band) has been used in the Bay Area of California (Cifelli et al. 2018; Willie et al. 2017), in Utah (Campbell and Steenburgh 2014), and in the Alps (Delrieu et al. 2009; Germann et al. 2006). The Swiss have developed operational methodology for using these observations in complex terrain, taking care to site the radars at high-elevation locations with clear views, to minimize sidelobe effects, and to correct for clutter (Germann et al. 2006). In Taiwan, vertical profile corrections were able to improve radar precipitation estimates where near-ground levels were blocked and only high-altitude signals were available.
(Wang et al. 2016). While limited access and funding will still cause problems, other locations could learn from these examples and conduct studies to improve processing algorithms for gap filling radars in locations where they can be feasibly deployed. Another approach to improve the use of radar data is through fusion with weather models. The assimilation of both radar-measured 3D reflectivity and wind velocities with an atmospheric model is an area of active research (Benjamin et al. 2016b; Sun 2005; Wang et al. 2013). However, application in complex terrain requires significant development (Tai et al. 2017).

Another example of the use of nontraditional data sources in precipitation mapping includes the use of ecological information. Maps of ecological communities may not report storm-specific precipitation amounts, but in remote regions, they provide high-resolution spatially detailed patterns of precipitation climatologies. Giambelluca et al. (2013) used both vegetation maps and high-resolution atmospheric model output to guide underlying patterns for precipitation across the mountains of Hawaii, and these climatological maps have been used to create many gridded precipitation products in near-inaccessible and unmonitored locations (Newman et al. 2019).

Soil moisture observations can identify areas where rain occurred recently but either are very localized (point measurements) or have large footprints (~40 km for satellites) and only see the very top layer of the soil, for example, Soil Moisture and Ocean Salinity (SMOS; Srivastava et al. 2015) and Soil Moisture Active Passive (SMAP; Chan et al. 2018). The Gravity Recovery and Climate Experiment (GRACE; and its follow-on mission), can identify total changes in water content at large spatial scales (~200,000 km²), which can benchmark modeled total changes in frozen and subsurface water storage (Chen et al. 2017), but GRACE has a large footprint and cannot identify from which water storage component (snow, ice, or groundwater) a storage change originated from. Alternative soil moisture measurements based on cosmic-ray neutron measurements (Zreda et al. 2008) or GPS interferometry (Larson et al. 2008) both offer intermediate spatial scales (tens of meters), and a large network of GPS installations has existed for decades.

**Snow measurements.** Of all the ancillary land surface data available, snow observations come closest to directly informing precipitation accuracy, particularly in areas where precipitation is difficult to directly observe. Snow water equivalent (SWE) observations are part of many in situ operational networks and have been used to assess total annual frozen precipitation as discussed above. Passive microwave observations of SWE from satellites are not robust for snow in the mountains, due to the large spatial footprint (~25 km), and the inability to sense deep (>20 cm) or wet snow (Dietz et al. 2012). Airborne gamma ray estimates of SWE are valid for shallow snow and flat terrain but are inaccurate in deep snow, forests, or complex terrain (Glynn et al. 1988). Thus, most spatially integrated assessments use other snow variables, such as snow depth and snow covered area, as described below.

**Snow disappearance date and the energy balance.** For decades, hydrologists have compared distributed snow simulations with satellite-based estimates of snow-covered area (SCA) or snow cover extent (Shamir and Georgakakos 2006), which provides a time-integrated assessment of combined model performance for both snow accumulation and melt with about 500 m (MODIS) or 30 m (Landsat) spatial resolutions, with new satellites offering the potential for 3 m (Planet) daily observations. A number of studies have utilized the date of snow disappearance to determine how much snow must have fallen at a location in order for snow to disappear on an observed date given modeled melt rates, often termed SWE reconstruction (Bair et al. 2016; Cline et al. 1998; Livneh et al. 2014; Molotch and Margulis 2008; Raleigh and Lundquist 2012). Many of these utilize Bayesian methodology and changing precipitation weights (Margulis et al. 2015, 2016), and some have directly used the reconstructed product to look at snowfall patterns in space and time (Huning and Margulis 2017, 2018). Wrzesien et al. (2017) found that the Margulis et al. (2016) data set compared favorably with WRF-based snow model simulations over the California Sierra Nevada. These techniques depend on accurately modeling the snow energy balance controlling melt rates. Thus, efforts to improve energy flux observations (e.g., snow albedo, snow surface temperature, radiation) will translate to improved retrospective snowfall estimation.

**Snow depth.** Snow depth observations can be obtained across large areas and at high resolutions (1–3 m spatial footprint) by repeat lidar (Painter et al. 2016) or satellite stereo photogrammetry techniques (Shean et al. 2016), but the direct application of these measurements to improving precipitation estimates has only just begun to be explored (Henn et al. 2016, 2018c; Mott et al. 2014). Snow depth is subject to localized effects such as preferential deposition, settling, and wind redistribution but can be aggregated to coarser resolutions for comparison with atmospheric...
models (Mott and Lehning 2010; Mott et al. 2011). Most atmospheric and hydrologic models use snow water equivalent as a state variable, which is related to snow depth via bulk density. The density of new snowfall is highly uncertain (Wayand et al. 2017), making density estimates one of the larger sources of error in deriving SWE from lidar measurements (Raleigh and Small 2017). Snow compaction and evolution through time are better understood, resulting in more confidence in density estimates in the spring near peak snow accumulation, making seasonal assessments more accurate than storm-specific assessments. Impacts of snow evolution due to rain on snow (which may increase both water content and density of the snowpack but generally to an unknown extent) are also a source of uncertainty.

Combining disparate data sources through partnerships across disciplines. We propose that two interrelated barriers, time scales and analysis tools, have impeded the rate of crossing historical disciplinary boundaries. Vegetation maps reveal finescale spatial patterns but only long-timescale (e.g., multidecadal average) precipitation patterns. While streamflow gives a solid estimate of total rainfall over a catchment, it is at the cost of smoother spatial and temporal resolutions (Kretzschmar et al. 2016). This smoothing in space and time makes streamflow and snow observations useful for evaluating mesoscale model output at an annual time scale, which fits well within a water resources perspective or a climate modeling development framework. Shorter time scales related to floods, particularly those due to convective precipitation events, are more difficult to assess using these techniques. Overall, both point observations and models struggle to accurately represent convective precipitation, and published research to date does not clearly indicate whether model simulations or gridded precipitation products are currently superior in this area. The spatial extent of convective precipitation can sometimes be identified more reliably with streamflow measurements than rain gauges in the mountains (Lundquist et al. 2009), but the precipitation magnitudes are hard to assess due to unknown water storage and evapotranspiration on the landscape over short time scales. Satellite data reveal convective patterns although amounts are biased (Rasmussen et al. 2013). Convolution-resolving models (4 km or finer) currently look like a viable path forward, although reproducing specific convective storms remains a challenge (Rasmussen et al. 2019).

Analysis tools also contrast between the two disciplines. Stochastic parameter selection and other statistical methods are common within hydrology and other land surface sciences, whereas deterministic methods dominate within atmospheric science—due in part to the computational costs involved. Hydrologists frequently run large model ensembles, ranging from one model with many different choices of parameter values, to a suite of models with different model choices in the equations used, often termed structural uncertainty (Clark et al. 2011). Comparisons of multiple long time series with observations are used to select choices that minimize errors. Numerical weather prediction modeling systems also often use ensembles (Hacker et al. 2011; Knievel et al. 2017), although ensembles using perturbed parameters (Greybush et al. 2017; Jankov et al. 2019) are rarer than those generated through perturbed initial conditions. There is invariably a trade-off between the size of the ensemble and the resolution and sophistication of the model (Ferro et al. 2012; Gowan et al. 2018). In general, testing uncertainty in parameter values or in both parameters and structure is more frequent in hydrology, while testing uncertainty in model structure alone is more frequent in atmospheric sciences.

These different approaches impact the necessary duration of measurements and of desired characteristics of mesoscale model output. Statistical methods become more reliable as the amount of training data (e.g., duration of time series) increases. Thus, many flood forecasters use gridded gauge data up until the time of the forecast to provide a long time period for model calibration and spinup, and an ensemble of weather model output for the future forecast (Pagano et al. 2014). Most research hydrologists prefer reanalysis datasets to provide meteorological forcings for their models because these models are run with a constant configuration over a long period of time. These allow hydrologists to calibrate unknown parameters in a hydrologic model (e.g., those related to subsurface storage and conductivity) and accommodate for biases in meteorological input (critical for flood forecasting or water resources operations).

In contrast, the increased value of physically based formulas is much less data volume dependent. Operational numerical weather prediction models change frequently to include the latest improvements in the modeling prediction system, which may be based on one short-term field campaign. However, this near-constant evolution makes them problematic for hydrologists since historic model calibration may introduce compensating biases that lead to the correct answer for the wrong reason (Kirchner 2006).

There are benefits to both stochastic and deterministic approaches, and both communities would benefit
from better sharing and integrating these techniques. For example, fully coupled land–atmosphere models currently lag behind the most promising hydrologic modeling results in mountainous terrain (Clark et al. 2015). This is likely the result of the coupled models being recently developed, often at coarser scales and with fewer processes represented, as well as the relative lack of model parameter estimation implemented in these studies. One path forward requires coupled models developed by actively engaged scientists from multiple disciplines and multiple communities because components that historically were unimportant to one or both disciplines will emerge as important in the coupled system. As atmospheric model output is more routinely adopted as hydrologic model input, as will happen with work currently underway with both the U.S. National Water Model (http://water.noaa.gov/about/nwm) and the European Flood Awareness System (www.efas.eu/about-efas.html), greater feedback and assessments will accumulate, which will improve both fields.

Strategically designing field campaigns requires a strong interdisciplinary approach in the earliest stages of planning. While atmospheric events pass within a week, the hydrologic cycle needs to be evaluated over the course of a year at minimum. Recent campaigns such as NOAA’s Hydrometeorology Testbed (HMT) in California (Ralph et al. 2005), the Olympic Mountain Experiment (OLYMPEX) in Washington (Houze et al. 2017), and the Integrated Precipitation and Hydrology Experiment (IPHEx) in North Carolina (Barros et al. 2014; Tao et al. 2016) provide examples of how instrument deployments have been balanced across longer and shorter durations to attempt to meet both atmospheric and hydrologic objectives. Further integration across time scales could be achieved by conducting shorter-duration atmospheric-focused campaigns at field locations that already have a long history of ground-based observations, for example, Long-Term Ecological Research (LTER; Kratz et al. 2003), Critical Zone Observatory (CZO; Lin and Hopmans 2011), or USDA research watersheds (Renard et al. 2008).

Despite efforts to encourage interdisciplinary work (Dirmeyer et al. 2015; National Research Council 1991), strong disciplinary boundaries still exist within universities and many funding agencies. The community should continue to advocate for interdisciplinary graduate education and for projects that combine knowledge and expertise across fields to advance these problems.

CONCLUSIONS. Our computational abilities have advanced more than our observational capabilities. Direct measurements of precipitation in complex terrain continue to be problematic. Even a concerted effort to invest financially in a network of mountain precipitation gauges may not improve our total gridded precipitation estimates significantly due to issues of access, property ownership, representativeness, and maintenance. Multiple types of measurements, including gauges, radars, satellites, streamflow observations, and snow observations (Fig. 3), must be brought to bear to assess our model output of mountain precipitation, and the combined modeling system must strive for excellence in all components relevant to those properties we can reliably measure (e.g., snow and streamflow). In the future, better coordination across disciplines could allow for improved coordination between the communities to improve available datasets available for hydrologic applications (Table 1).

### Table 1. Summary advice.

<table>
<thead>
<tr>
<th>Advice for atmospheric scientists</th>
<th>Advice for land surface scientists</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Do not presume that your model’s inability to match a gridded dataset means that it is wrong (Hughes et al. 2019; Liu et al. 2017)</td>
<td>• High-resolution atmospheric model output is likely superior input to your hydrologic model than distributed gauge data</td>
</tr>
<tr>
<td>• Investigate multiple sources of data, as well as many different types of storms, to make decisions between different model setups and parameterizations</td>
<td>• Atmospheric models can output amounts of mixed-phase precipitation directly, often with better success than empirical methods based on surface observations of air temperature and humidity</td>
</tr>
<tr>
<td>• Consult hydrologists and snow scientists about the sources of uncertainty in streamflow, soil moisture, and snow measurements (useful, independent data)</td>
<td>• Not all simulations are created equal—befriend some modelers and ask questions about the parameterization schemes, resolution, nesting, etc.</td>
</tr>
<tr>
<td>• Consult those working in more application-oriented fields to understand what they really need (e.g., getting precipitation on the correct side of an orographic barrier is important)</td>
<td>• Work with meteorologists to better understand how atmospheric variables that influence hydrology may be discontinuously distributed in space and time (i.e., stationarity or linearity is often assumed, but may not hold in reality)</td>
</tr>
</tbody>
</table>
Realizing that our current modeling capabilities may now be surpassing our observational abilities in the realm of mountain precipitation speaks well to human ingenuity and scientific improvements over recent decades. We must take this realization as inspiration to now maintain and improve our observational networks and our capabilities in fully utilizing them in order to continue to improve our abilities to understand, model, and forecast our mountain water supplies, transportation hazards, and the elusive deep powder.

ACKNOWLEDGMENTS. We thank Lucas Harris, Chris Milly, Haonan Chen, Rob Cifelli, and two anonymous reviewers for their reviews of the manuscript. We also thank Catherine Raphael for her assistance in graphical design. The National Center for Atmospheric Research is sponsored by the National Science Foundation. J. Lundquist was partially supported by NASA Grant NNX17AL59G.

REFERENCES


——, K. L. Rasmussen, M. D. Zuluaga, and S. R. Brodzik, 2015: The variable nature of convection in the tropics and subtropics: A legacy of 16 years of


——, J. D. Lundquist, and B. Henn, 2019: Dynamical downscaling improves upon gridded precipitation products in the Sierra Nevada, California. *Climate Dyn.*, https://doi.org/10.1007/s00382-017-3631-z, in press.


Kuczera, G., D. Kavetski, S. Franks, and M. Thyer, 2006: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising


Minder, J. R., and D. E. Kingsmill, 2013: Mesoscale variations of the atmospheric snow line over the northern Sierra Nevada: Multiyear statistics, case


Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations

OAR authors.


