Observed humidity trends in dry regions contradict climate models

Isla R. Simpson[a,1] (iD), Karen A. McKinnon[b,c,d] (iD), Daniel Kennedy[a,e], David M. Lawrence[a], Flavio Lehner[a,f,g] (iD), and Richard Seager[h] (iD)

Arid and semi-arid regions of the world are particularly vulnerable to greenhouse gas–driven hydroclimate change. Climate models are our primary tool for projecting the future hydroclimate that society in these regions must adapt to, but here, we present a concerning discrepancy between observed and model-based historical hydroclimate trends. Over the arid/semi-arid regions of the world, the predominant signal in all model simulations is an increase in atmospheric water vapor, on average, over the last four decades, in association with the increased water vapor–holding capacity of a warmer atmosphere. In observations, this increase in atmospheric water vapor has not happened, suggesting that the availability of moisture to satisfy the increased atmospheric demand is lower in reality than in models in arid/semi-arid regions. This discrepancy is most clear in locations that are arid/semi-arid year round, but it is also apparent in more humid regions during the most arid months of the year. It indicates a major gap in our understanding and modeling capabilities which could have severe implications for hydroclimate projections, including fire hazard, moving forward.

climate change | modeling | humidity | hydroclimate

One of the most pressing issues facing society and ecosystems as the planet warms is the impact of a changing hydroclimate and its associated effects on drought, wildfire, and heat extremes. This is particularly true in arid and semi-arid regions where water resources are limited, and wildfire and heat extremes are already a significant threat. The US Southwest is a particularly clear example, having been dominated by drought for the last 20 y with recent extreme conditions that led to unprecedented water shortages in the Colorado River (1) and extreme wildfire seasons (2) that have almost certainly been exacerbated by greenhouse gas–driven warming and aridification (1, 3, 4). Burned forest area in the Southwest is highly correlated with vapor pressure deficit, the difference between saturation and actual vapor pressure (2, 5, 6), so drought impacts can stem from changes in precipitation, temperature, and atmospheric humidity. What kind of future hydroclimate extremes should the Southwest, and regions like it, be preparing for?

Climate models, which simulate the complex interacting processes that govern the hydroclimate, are an important tool for answering this question. A challenge is that many of the relevant processes or quantities such as evapotranspiration, root zone soil moisture, and plant physiological changes have not been observed on the global scale or on the multi-decadal timescales over which the planet has been changing, to evaluate our models. We do, however, have a reasonably complete network of station-based near-surface atmospheric humidity measurements as well as reanalysis-based estimates of atmospheric water vapor (7–9). In some sense, near-surface water vapor should act as an integrator of how processes that are of relevance to the hydroclimate are evolving, so a discrepancy in atmospheric water vapor trends between models and observations would be indicative of something being wrong in our model representation of processes that are of relevance to the hydroclimate, assuming atmospheric water vapor observations can be trusted.

Our expectations for how atmospheric water vapor should change with warming are based on the Clausius–Clapeyron relationship, which links saturation vapor pressure (SVP) to temperature. At the temperature range found at the Earth's surface, it roughly predicts a rise in SVP of about 7% per Kelvin of warming. Whether atmospheric water vapor would rise at this rate, to maintain a constant relative humidity, depends on the availability of moisture from transport or from evapotranspiration. In the global and total column average, atmospheric water vapor in climate models and observations does tend to rise at approximately this rate (10, 11). Regionally near the surface over land, it is less clear that Clausius–Clapeyron scaling will hold, for a number of reasons. Atmospheric circulation change can lead to altered moisture transports. The land also warms more

Significance

Water vapor in the atmosphere is expected to rise with warming because a warmer atmosphere can hold more moisture. However, over the last four decades, near-surface water vapor has not increased over arid and semi-arid regions. This is contrary to all climate model simulations in which it rises at a rate close to theoretical expectations, even over dry regions. This may indicate a major model misrepresentation of hydroclimate-related processes; models increase water vapor to satisfy the increased atmospheric demand, while this has not happened in reality. Given close links between water vapor and wildfire, ecosystem functioning, and temperature extremes, this issue must be resolved in order to provide more reliable climate projections for arid and semi-arid regions of the world.

than the oceans (12, 13) so the increased water vapor transport from the oceans may be insufficient to keep up with Clausius Clapyeron scaling over land (14). Furthermore, over land, limited surface water availability, or changes to surface water availability (either from plants or bare ground), can limit the extent to which evapotranspiration can increase to satisfy the increased atmospheric demand. Despite this, while model projections for the end of the 21st century do exhibit a slight decline in relative humidity over land, they still suggest that atmospheric water vapor should increase (15).

Prior studies on historical atmospheric water vapor trends over land found rising specific humidity in general but, in association with the hiatus in the warming of the oceans at the beginning of the 21st century, specific humidity remained fairly constant over the 2000s with a concomitant relative humidity decline (16–18), which was likely exacerbated by continued warming of extreme warm temperatures over land (19). A number of studies have compared modelled and observed relative humidity trends over large geographical regions and found that the observed relative humidity decline has been greater than typically found in models (17, 20–22). Here, with an observational record that extends beyond the warming hiatus of the early 21st century, and the latest generation of global climate models, we further scrutinize the long-term humidity trends in models in comparison to observations and will demonstrate a major discrepancy in their behavior that has strong ties to climatological aridity and is most apparent in arid and semi-arid locations.

A Case Study: The Southwest United States

We first demonstrate the discrepancy in humidity trends using the US Southwest, which has a dense network of in-situ humidity measurements (7) (*Materials and Methods*). Annual mean vapor pressure (VP) trends from 1980 to 2020 in the HadISDH (8) stations and the ERA5 reanalysis (9) (Fig. 1 *A* and *B*) agree and demonstrate a substantial decline over the Southwest (23). While agreement between ERA5 and the stations is not surprising, given that ERA5 assimilates station-based humidity measurements, it is reassuring and indicates that the ERA5 trends are not the spurious result of changes in observed data streams, or issues in the underlying forecast model (see the *Materials and Methods* for a discussion of other reanalyses). This VP decline can be compared with the distribution of trends in CMIP6-era models in Fig. 1*C* (x-axis). The model members overwhelmingly suggest that Southwest VP should have increased; the 5th to 95th percentile range across the CMIP6 members is 0.14 hPa/41 y to 1.04 hPa/41 y and the minimum modeled trend of −0.19 hPa/41 y falls drastically short of the ERA5/ISD/HadISDH declines of −0.81/−0.70/−0.58 hPa/41 y. Simulations with prescribed observation-based sea surface temperatures (SSTs), which are thought to have driven a precipitation decline over the region (24, 25), also cannot reproduce the observed VP decline see the CESM2 Global Ocean Global Atmosphere (GOGA2) points in Fig. 1*C*. These conclusions are not strongly affected by ending the trend in the 2020 drought year (Fig. 1*C*, open gray diamond).

For SVP, which is only dependent on temperature, the observed trends lie within the model distribution (y-axis of Fig. 1*C*), albeit with the models tending to warm more (and show a greater increase in SVP) on average. The difference between the SVP and the actual VP (the vapor pressure deficit, VPD) is a highly relevant quantity for wildfire (2, 5, 26). It has already been recognized (27, 28) that VPD has increased over the Southwest with contributions from both increasing SVP and decreasing

actual VP, and Fig. 1*C* (gray contours) indicates that, while the model trends in VPD encompass the observed, they do so for different reasons: A greater increase in modeled SVP (more warming) compensates for not simulating a decline in actual VP.

Time series of Southwest VP anomalies also compare well between ERA5 and the stations (ISD and HadISDH) (Fig. 1*D*). The observed decline in VP and the difference from CMIP6 is relatively homogeneous in time (*SI Appendix*, Fig. S7) and is not dominated by an episodic jump (or jumps) in the record. This, together with prior literature on the fidelity of station-based humidity measurements (see *Materials and Methods* for an in-depth discussion), suggests that the decline is not a spurious result of changes in observing practices. Overall, this leads us to believe that the station-based observations represent a true trend that is also well captured by the assimilation in ERA5.

The observed decline in humidity alongside rising temperature leads to a substantial relative humidity decline (Fig. 1*E*) that is also not well captured by the models. The discrepancy in specific humidity trends between ERA5 and the models maximizes at the surface but is apparent throughout the lower troposphere (Fig. 1*F*) and is relatively uniform throughout the year when considering the trend as a percentage of the monthly climatology (Fig. 1*G*). However, in an absolute sense, the difference in VP trends maximizes in the summer when the climatological VP is highest (*SI Appendix*, Fig. S8).

Precipitation has declined in the Southwest during the study period, which we expect would cause a decline in VP. Trends in VP and precipitation are correlated across model members in both CMIP6 and in LENS2 (Fig. 2*A*). In LENS2, the difference in trends between the members is entirely due to internal variability and since the LENS2 spread in precipitation trends and VP trends is almost as large as that in CMIP6, it suggests that internal variability is a dominant contributor to spread in CMIP6. While the observed decline in precipitation is large, it is not outside of the range of modelled trends. In order to account for the influence of precipitation trends, we fit a regression line across CMIP6 members between VP and precipitation trends (*Materials and Methods*). The observed decline in VP is much larger than would be expected after conditioning on the precipitation trends in this way and lies at much lower values than any of the CMIP6 residuals from the regression fit (Fig. 2*A* and see *Materials and Methods*).

Considering the shorter 1980 to 2014 period, over which we can examine prescribed SST experiments from other CMIP6 models (AMIP6, see *Materials and Methods*), it is clear that prescribing observed SSTs leads to lower VP trends on average (*SI Appendix*, Fig. S9*A*). This is because observed SSTs force a greater precipitation decline (*SI Appendix*, Fig. S9*B*). However, while there is a greater overlap between AMIP6 and observations, those models that exhibit a VP decline as low as observed do so because they exhibit a precipitation decline that is much greater than observed. The observed VP decline is, therefore, also inconsistent with the behavior over this shorter time period of this broader range of prescribed SST simulations.

Overall, the observed decline in near-surface vapor pressure is much larger than what is expected based on the models, even after accounting for the relatively large observed Southwest precipitation decline. Fig. 2*A* suggests three possible origins of this issue: 1) The slope of the relationship between the VP trend and precipitation trend could be too shallow; 2) the forced, thermodynamic, rise in atmospheric water vapor which shifts the entire distribution in Fig. 2*A* toward positive VP values could be too large in the models; or 3) models could be deficient
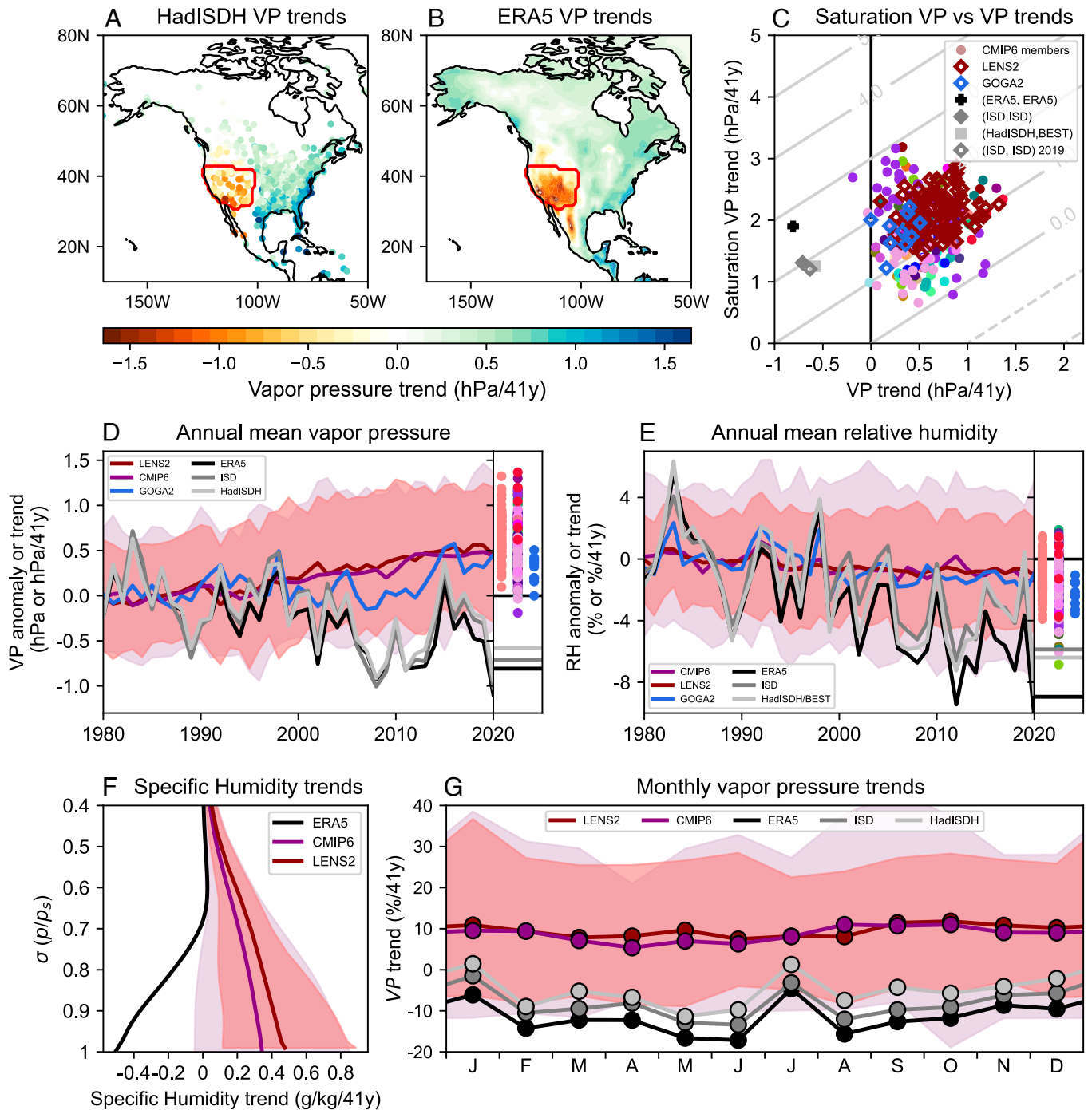
Fig. 1. Annual mean 1980 to 2020 water vapor trends over the US Southwest. (*A* and *B*) Vapor pressure trends in HadISDH stations and ERA5 (*Materials and Methods*), respectively with red depicting the six state region (California, Nevada, Arizona, New Mexico, Utah, and Colorado, referred to as the Southwest) used in other panels (see *SI Appendix*, Fig. S2A for the trends in the other ISD dataset). (*C*) Southwest saturation vapor pressure trend versus actual vapor pressure trend with CMIP6 members depicted by colored dots (different colors for each model), LENS2 members depicted by maroon diamonds, and GOGA2 members depicted by blue diamonds. Combinations of observation-based data are shown in black and gray with the legend listing the dataset used for humidity first and for temperature second. The gray diagonal lines depict the trends in VPD in hPa/41 y. The open gray diamond shows the ISD trends ending in 2019 but scaled by 41 y. (*D*) Time series of annual mean vapor pressure. For CMIP6, the shading depicts the 2.5th to 97.5th percentile range of all members from all models, and for LENS2, the shading depicts the 2.5th to 97.5th percentile range of the members. The *Right* portion of the panel depicts the overall trend for individual members. (*E*) is as (*D*) but for relative humidity. (*F*) Vertical profiles (on sigma levels, p/p_s, with the interpolation to sigma levels performed at each grid point prior to taking the Southwest average) of specific humidity trends with shading for CMIP6 and LENS2 indicating the minimum to maximum member range. (*G*) 1980 to 2020 vapor pressure trends for individual months of the year expressed as a percentage of the 1980 to 1990 climatology for that month of the year with the shading showing the minimum to maximum range across model members.

in their representation of the natural variability of atmospheric water vapor that is unrelated to precipitation, i.e., the distribution of points around the regression line in Fig. 2*A* should perhaps

be broader in the vapor pressure trend direction. In detrended, interannual variability, there is a similar slope of the relationship between VP and precipitation to that seen for long-term trends
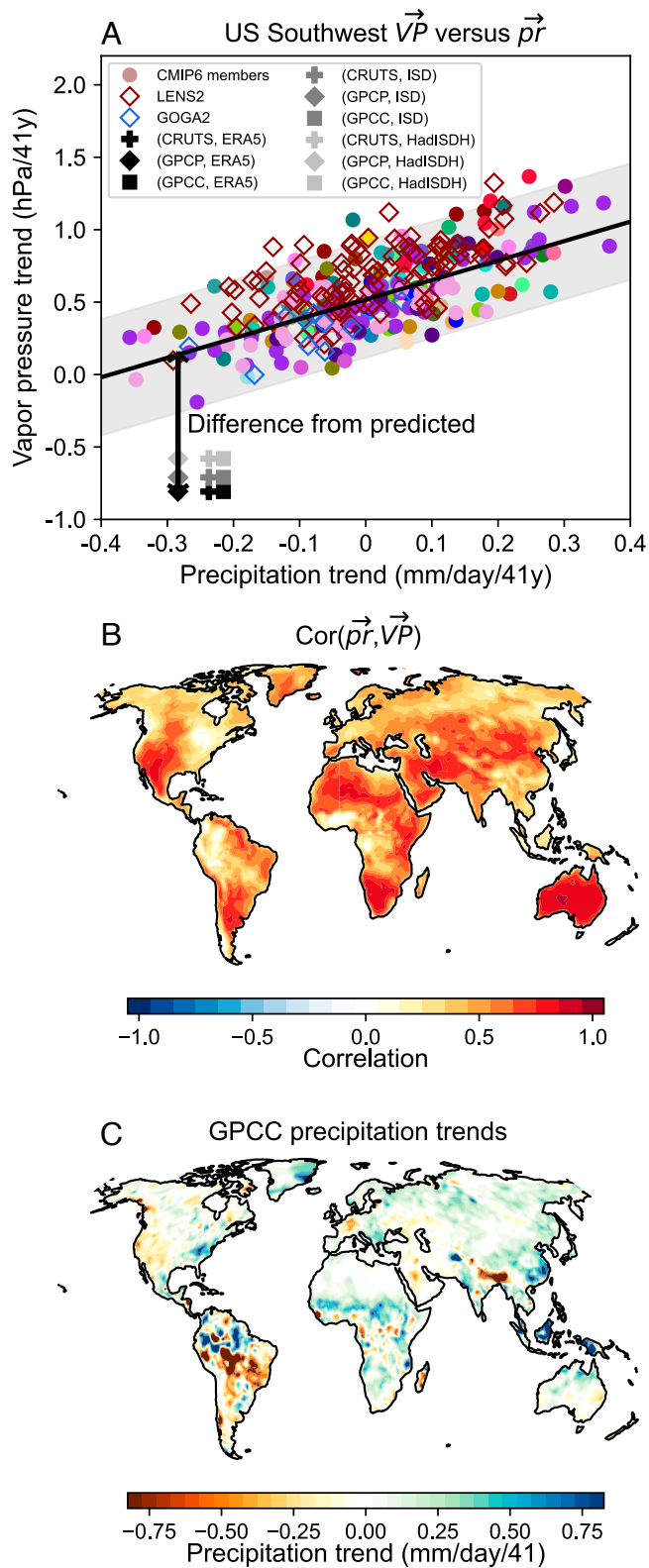
Fig. 2. The relationship between 1980 to 2020 annual mean vapor pressure trend $\overrightarrow{VP}$ and precipitation trend $\overrightarrow{pr}$. (A) Vapor pressure trend versus precipitation trend averaged over the US Southwest for the CMIP6 members (colored dots, with each color representing a different model with the number of members summarized in *SI Appendix*, Table S1), LENS2 (red diamonds), GOGA2 (blue diamonds), and various observation-based datasets (gray and black diamonds, crosses, and squares). The diagonal line shows the least squares regression fit to the CMIP6 members, and the gray shaded range shows ±2σ around this regression line where σ is the SD of the residuals of the individual CMIP6 members from the regression fit. (B) The correlation between precipitation trend and vapor pressure trend using the CMIP6 members. (C) Observation-based precipitation trends using GPCC data.

in Fig. 2*A* and observations agree well with the CMIP models in this metric (*SI Appendix,* Fig. S10 *A–C*). This provides hints that (1) is not the problem, but in the following, we will further exploit the near-surface humidity trends globally to confirm this and demonstrate that (3) is also not the primary issue, but rather (2) is the most likely candidate.

Global Near-Surface Vapor Pressure Trends

Globally, there is good agreement between ERA5 and the station-based VP trends (Fig. 3 *B* and *C* and *SI Appendix,* Figs. S5 and S6). Aside from the US Southwest, there are a number of regions where the 1980 to 2020 ERA5 VP trend lies outside of the CMIP6 member distribution of 1980 to 2020 trends, and these are all regions where the VP trends are negative or near zero (Fig. 3*C*): Northern Argentina, South Africa, South Sudan, Somalia, Spain and Morocco, Iran, Kazakhstan, Mongolia, and East Australia. Precipitation has not declined at all these locations (Fig. 2*C*) and since precipitation and VP trends tend to be correlated (Fig. 2*B*), we incorporate precipitation trends into the analysis.

Fig. 3*D* shows the difference between ERA5 VP trends and the VP trends that would be predicted based on observed precipitation trends and the CMIP6 relationship between VP and precipitation trend (See *Materials and Methods* and Fig. 2*A* for a demonstration of the approach, and *SI Appendix,* Fig. S11*B* for the actual VP trends that are predicted in this way). Many more regions now have ERA5 VP trends that are distinct from the model distribution, and these are almost all regions where the ERA5 trends, after accounting for precipitation trends, are relatively negative (except for Saudi Arabia and parts of India). The same is true in the CMIP5 models (*SI Appendix,* Fig. S20) and when considering the shorter 1980 to 2014 trends in both CMIP6 and AMIP6 (*SI Appendix,* Fig. S12). Importantly, the regions where we see this discrepancy are not randomly distributed but correspond to arid and semi-arid regions around the globe (Fig. 3*A* and see *Materials and Methods* for Aridity Zone definitions).

We formalize this conclusion by considering the land surface VP trends binned into equal-area bins according to the climatological Aridity Index and precipitation trends in Fig. 3 *E–G* (*Materials and Methods*). This makes it clear that the annual mean ERA5 VP trends are distinct from the modeled trends almost exclusively within arid/semi-arid regions and the discrepancy occurs regardless of the precipitation trend. Differences between modeled and observed VP trends could be caused by differences in temperature trends but the latter are relatively small in this aridity–precipitation trend space (*SI Appendix,* Fig. S14). Furthermore, the same conclusions hold true when we account for the temperature trend over land as shown when binned specific humidity trends are expressed per Kelvin of land area-averaged (excluding Antarctica) temperature trend (Fig. 3 *H–J*). In the CMIP6 ensemble mean, the specific humidity trends are well aligned with Clausius–Clapeyron scaling (∼7%/K) across the full range of Aridity Indices, with deviations to lower or higher values depending on the precipitation trend (Fig. 3*I*). In contrast, the specific humidity trends in ERA5 in arid and semi-arid regions are much lower than Clausius–Clapeyron scaling and even negative where precipitation declines (Fig. 3*H*) in regions like the US Southwest. Overall, CMIP6 models suggest that the historical humidity trends in arid/semi-arid regions should have been close to Clausius–Clapeyron scaling on average, but this is at odds with reality.

The close link between climatological aridity and the discrepancy in annual mean humidity trends raises the question, does
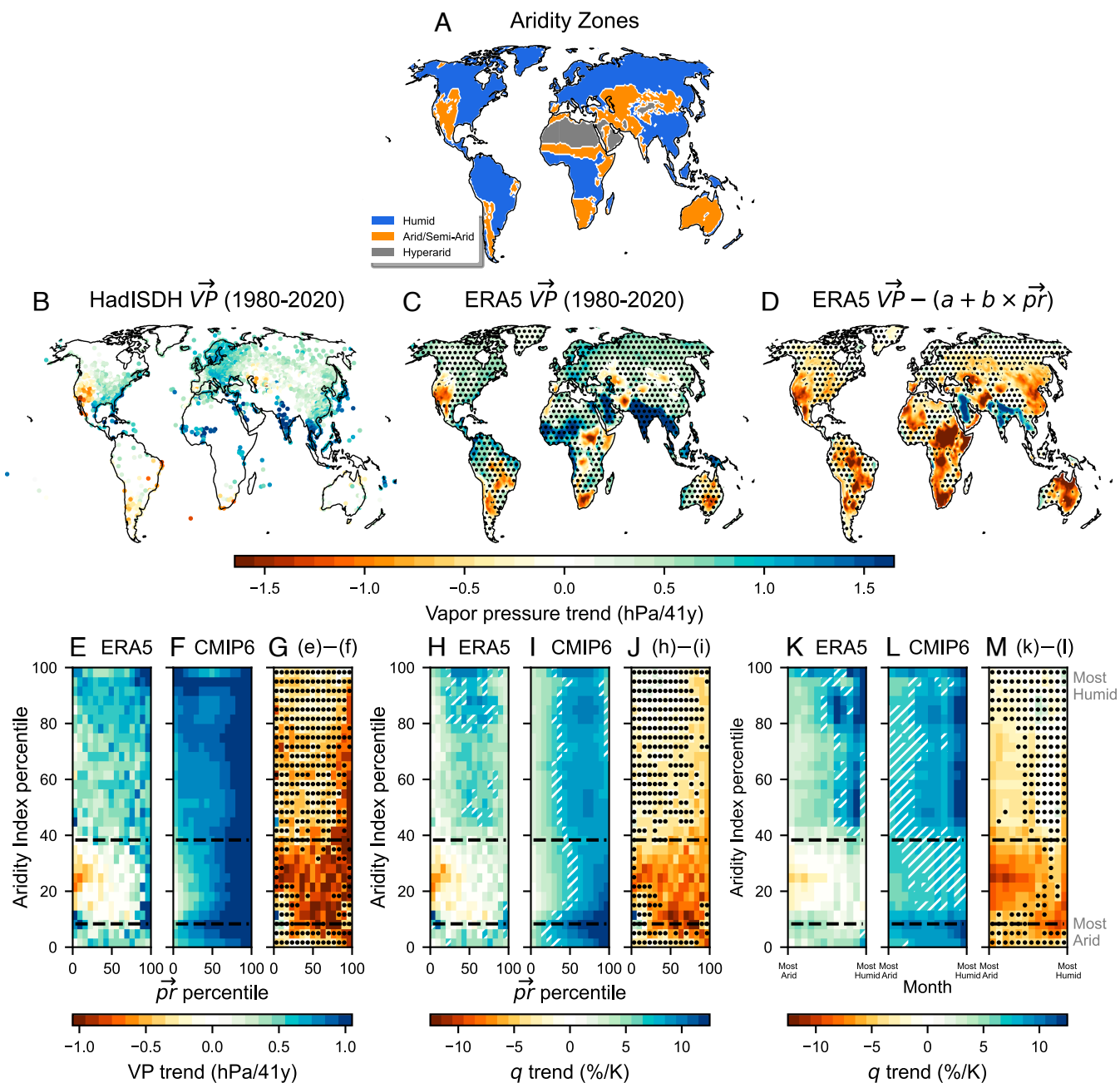
Fig. 3. The global discrepancy in 1980 to 2020 water vapor trends. (*A*) indicates the land regions classified as humid (blue), arid/semi-arid (orange), and hyperarid (gray). (*B*) Annual mean 1980 to 2020 VP trends in the HadISDH station data. (*C*) as (*B*) but for ERA5, and the stippling indicates where the ERA5 VP trend lies within the range of the CMIP6 members. (*D*) The difference between the ERA5 VP trend ($\overrightarrow{VP}$) and that predicted based on the precipitation trend ($\overrightarrow{pr}$) using the linear relationship $\overrightarrow{VP} = a + b \times \overrightarrow{pr}$ where a and b are determined for each grid point using the CMIP6 models and GPCC for observed $\overrightarrow{pr}$ (see *SI Appendix,* Fig. S11 *F* and *G* for alternatives). Stippling shows regions where the observed VP trend lies within the $\pm 2\sigma$ range of CMIP6 residuals around the $\overrightarrow{VP}$-$\overrightarrow{pr}$ relationship. (*E*) VP trends across land regions binned into 30 equal-area bins according to the annual mean Aridity Index and then binned into 15 equal-area bins according to the annual mean precipitation trend using GPCC (*Materials and Methods* and *SI Appendix,* Fig. S13 for alternative precipitation observations). Arid/semi-arid zones sit within the black dashed lines. (*F*) is as (*E*) but for the CMIP6 ensemble mean using CMIP6 precipitation trends for the bins. (*G*) is as (*E*) but for the difference between ERA5 and the CMIP6 ensemble mean, with dots depicting where ERA5 lies within the CMIP6 model and ensemble member range. (*H–J*) are as (*E–G*) but for specific humidity trend expressed as a percentage of the 1980 to 1990 climatology per trend in land area average temperature (excluding Antarctica). (*K–M*) are as (*H–J*) showing specific humidity trend normalized by land area averaged temperature trend, but trends for each month of the year are used and rather than binning according to precipitation trend in the x-axis, the months of the year are ordered from most arid on the left to most humid on the right (based on the climatological monthly Aridity Index) (*Materials and Methods*). White hatching in panels (*H*), (*I*), (*K*), and (*L*) indicates bins where the specific humidity trend lies between 6.5 and 7.5% per K i.e., close to Clausius–Clapeyron scaling.

a similar discrepancy occur in more humid regions but during the most arid times of the year? In Fig. 3 *K–M*, monthly specific humidity trends normalized by temperature trend are shown. The months of the year are ordered from most arid to most humid at each location and depicted on the x-axis and all locations are

then binned into 30 equal-area bins according to the annual mean Aridity Index on the y-axis (*Materials and Methods*). The average precipitation trends in each bin are relatively small (*SI Appendix,* Fig. S15 *F* and *P*). While the model discrepancy is largest in Arid/Semi-Arid regions, and at the most arid time

of the year in those regions, the humid regions also show a significantly lower specific humidity trend on average but only during their most arid months of the year (Fig. 3M). During these months, the Aridity Index enters the realms of the aridity indices that arid/semi-arid regions see year round (*SI Appendix*, Fig. S15A), and these tend to be the hottest months with the greatest climatological specific humidity (*SI Appendix*, Fig. S15 B and D). Where and when models overestimate the rise in atmospheric humidity is, therefore, clearly linked to climatological aridity, both spatially and seasonally.

Focusing now exclusively on arid/semi-arid regions, as defined by the annual mean Aridity Index, we present the relationship between specific humidity trends and precipitations trends in Fig. 4A for the CMIP6 models compared to observations (See *SI Appendix*, Fig. S16 for a LENS2 version of this figure and *SI Appendix*, Fig. S20 for a version with CMIP5). In Fig. 2A, we examined the relationship between humidity trends and precipitation trends across model members in one location, but now we are examining this relationship across space in each member. A version of this figure for VP and not normalized by temperature trend, which can be more easily compared to Fig. 2A, is shown in *SI Appendix*, Fig. S17. This reveals a similar relationship across spatial grid points to that across model members in the Southwest, which is not too surprising since most arid/semi-arid regions exhibit a VP vs precipitation relationship with a similar slope to that in the Southwest (*SI Appendix*, Fig. S10D). But since we are considering the relationship across space, we can now examine it in the observations as well. Fig. 4 indicates that the primary difference between the models and the observation-based data is that the model-based specific humidity trends are offset to higher values at all precipitation trends. While the model-based specific humidity trends are close to 7 %/K in regions with near zero precipitation trend, the observation-based specific humidity trends are close to zero. Over the 1980 to 2014 period when the AMIP6 simulations can be examined, the time evolution of SSTs does increase the likelihood of negative specific humidity trends slightly, but the same overall conclusion can be drawn with prescribed observation-based SSTs, i.e., the relationship between precipitation trends and VP trends is similar between models and observations, but the modelled VP trends are offset to positive values, while the observed trends are centered around zero (*SI Appendix*, Fig. S18).

There has essentially been no increase in water vapor over arid/semi-arid regions in the observation-based data, while there has been an increase that is close to Clausius–Clapeyron scaling in the models. This conclusion that we draw from the model ensemble mean is also valid for individual members as demonstrated by the probability distributions of specific humidity trends in Fig. 4 B and C. Conditioning on locations where the precipitation trend is low (between −0.05 and +0.05 mm/d/41 y), the difference in behavior between models and observations is dramatic (Fig. 4C); in ERA5, around 51.5% of the arid/semi-arid area with low precipitation trends exhibits a negative specific humidity trend, in contrast to the maximum area of 10.4% in CMIP6 and only 5.0% in LENS2 (*SI Appendix*, Fig. S16). There are also indications that the distribution of specific humidity trends is broader in the observation-based data than in the models, but this appears to be a secondary contributor to the difference in the distributions compared to the large offset between them on the humidity axis.

In the context of the US Southwest and the relationship between VP and precipitation trends there (Fig. 2A), we raised three possibilities that could be contributing to the model-observation discrepancy: a difference in the slope of the relationship between humidity and precipitation trends; a difference in
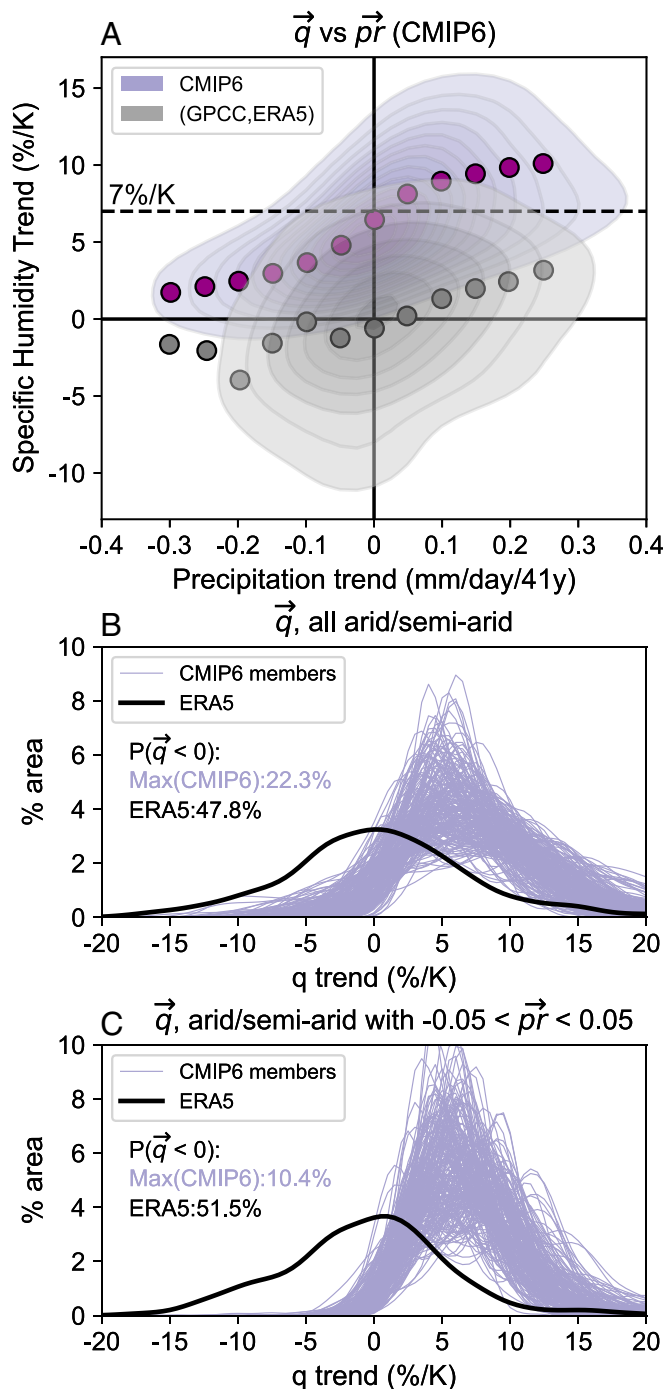

Fig. 4. Probabilities of specific humidity trends ($\vec{q}$) as a function of precipitation trends ($\vec{pr}$) in arid and semi-arid regions. (*A*) area weighted joint probability distribution of specific humidity trend (expressed as a percentage of the 1980 to 1990 climatology and normalized by the land area average temperature trend, excluding Antarctica) versus precipitation trend for arid/semi-arid regions using a Gaussian kernel density estimate. Gray shows observation-based data (GPCC for *pr* and ERA5 for *q*, see *SI Appendix*, Fig. S11 for alternatives) and purple shows the ensemble mean of the first member for each CMIP6 model (probability distribution calculated for each model first and then averaged). Contours show the probability per 0.025 mm/d/41 y $\vec{pr}$ by 0.5 %/K $\vec{q}$ box with contours starting at 0.04% and increasing in increments of 0.02% and the circles show the area weighted average specific humidity trend for $\vec{pr}$ bins of width 0.05 mm/d/41 y. (*B*) Gaussian kernel density estimate of the % area of all arid/semi-arid regions exhibiting specific humidity trends using 0.5 %/K bins. Black shows ERA5 and purple lines show all members from all models for CMIP6. The text depicts the percent area exhibiting negative trends for ERA5 and the CMIP6 member with the maximum area of negative trends. (*C*) is as (*B*) but now the probability estimates are performed only on arid/semi-arid grid points that exhibit precipitation trends between −0.05 and 0.05 mm/d/41 y.

the thermodynamic rise in water vapor with warming; and, a difference in the variability in specific humidity trends that is unrelated to precipitation. Here, aggregating across all arid and semi-arid regions of the world that differ in their historical precipitation trends, we can show that the primary issue lies in the representation of the forced thermodynamic rise in atmospheric water vapor that is unrelated to precipitation, although there is also slightly more variability in the observed specific humidity trends that is unrelated to precipitation. In observations, specific humidity has not increased over arid/semi-arid regions of the world in contrast to models where it increases at a rate close to Clausius–Clapeyron scaling on average. In the annual mean, the agreement between modelled and observed trends is much closer in humid regions (*SI Appendix*, Fig. S19), but as we have already shown in Fig. 3*M*, even in humid regions, the most arid months of the year are associated with lower humidity increase in ERA5 than in the models.

A summary of the historical time evolution averaged over arid and semi-arid regions is shown in Fig. 5 (see *SI Appendix*, Fig. S22 for all land regions and other aridity zones, and *SI Appendix*, Fig. S23 for comparison of ERA5 with gridded HadISDH). Near-surface air temperature has warmed over arid/semi-arid regions since 1980 (Fig. 5*A*) with relatively good agreement between models and observations. In association with this rise in temperature, atmospheric specific humidity increases in the models (Fig. 5*C*), and only a very small decline in relative humidity occurs (Fig. 5*D*). This is in sharp contrast to the observation-based record, where specific humidity has not risen and relative humidity has declined by about an order of magnitude more than the models on average. Prescribing observation-based SSTs does lower the increase in specific humidity during the hiatus period of the early 2000s when the warming of SSTs stalled on average (Fig. 5*B*) (11, 16, 17). Even so, specific humidity clearly still rises more, and relative humidity declines more, in the models compared to reality even when observation-based SSTs are prescribed. Further, the GOGA2 simulations exhibit a continued rise in atmospheric water vapor and become more aligned with the coupled models after 2014 as the ocean temperatures continue to rise.

Discussion

The observational record should be viewed with the caveat that it may be subject to errors since artifacts and artificial drifts can be introduced as instruments and observing practices change over time. For example, ISD stations tend to be located at airports, which could in theory exhibit different trend characteristics from the surrounding regions, and changes in instruments and perhaps station location have changed over time. However, we consider the evidence presented more fully in the *Station Data Fidelity* section in *Materials and Methods* to indicate that observational artifacts are unlikely to be the driver of this model versus observational discrepancy. Previous studies (29, 30) investigating station-based measurements of humidity from the United States have found minimal impacts of changes in instrumentation or urbanization on trends, and we have shown here that results are robust to different choices made in the selection and homogenization of station data through comparison of our own analysis of ISD with the homogenized HadISDH dataset. We further demonstrate in *SI Appendix*, Figs. S7 and S21 that the discrepancy is robust for a wide range of running vapor pressure trends with varying start points and lengths, which indicates considerable homogeneity in the bias in trends between observations and CMIP6 over the record. If observational artefacts are responsible, they would have to explain a) the global nature of the discrepancy, b) its localization
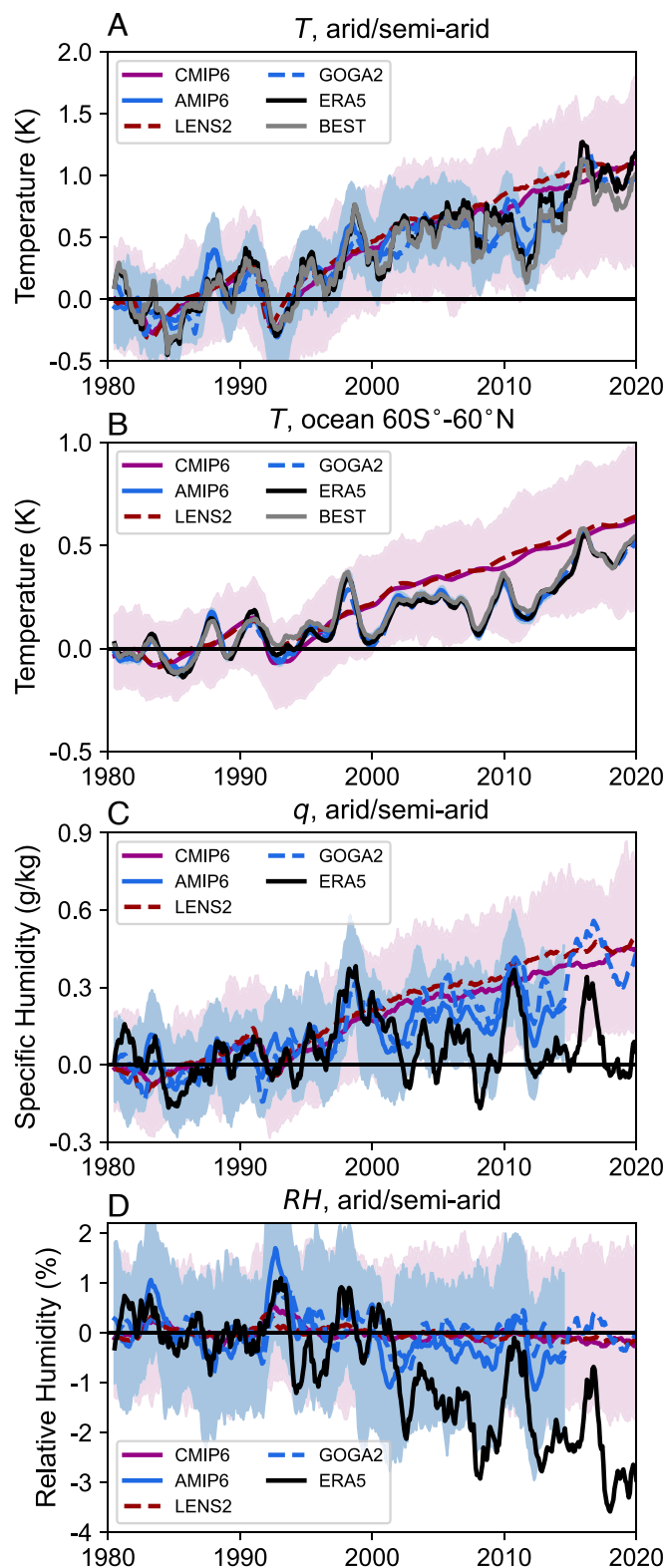
Fig. 5. (*A*) Area weighted average of 12-mo running mean near-surface air temperature (anomalies from the 1980 to 1990 average) over arid and semi-arid regions. Black shows ERA5, gray shows BEST, purple shows the CMIP6 ensemble mean, the pink range shows the 2.5th to 97.5th percentile range across all the members from all models for CMIP6, maroon dashed shows the LENS2 ensemble mean, blue solid shows the AMIP6 ensemble mean, the light blue shading shows the 2.5th to 97.5th percentile range across all AMIP members from all models and blue dashed shows the GOGA2 ensemble mean. (*B*) is as (*A*) but for the average of near-surface air temperature over oceans from 60°S to 60°N. (*C*) is as (*A*) but for near-surface specific humidity averaged over arid and semi-arid regions using ERA5 for observations which compares well with HadISDH (*SI Appendix*, Fig. S23), (*D*) is as (*C*) but for relative humidity averaged over arid and semi-arid regions.

to a particular climate type (arid/semi-arid regions or the most arid seasons in more humid regions), and c) the homogeneous nature of the trend discrepancy over time. We, therefore, argue that a more likely explanation is that some process is being misrepresented in models.

If a model issue is driving the discrepancy in humidity trends identified here, its origins are unclear, but we propose a few candidate mechanisms for further investigation: 1) too much water may be available from the land in models climatologically; 2) the land surface may have already dried out in response to warming more than the models simulate; 3) plant physiology may be responding differently to rising CO_2 in reality than in models (31); 4) runoff and/or its response to warming may be misrepresented in models, which could also be related to (1) and (2); and/or 5) there may be a global misrepresentation of the moisture transport response to warming. Efforts to distinguish between these and other possibilities must be guided by the fact that the issue is common to all models, the trend discrepancy is most severe in arid/semi-arid regions, and in more humid regions, it is only found in the most arid seasons. A simple explanation for the links with climatological aridity may be that the water limitations that are inherent to arid/semi-arid regions result in any small change in water availability having a larger impact on evapotranspiration and atmospheric humidity. If so, it is possible that the underlying model issue may apply more generally, but it is only in these water-limited regions and seasons that it has become detectable.

Some recent model-observation comparisons of evapotranspiration anomalies during dry conditions (32, 33) suggest that more water is available for evapotranspiration during drought in reality than in models, which runs counter to hypothesis (1). In the climatological average, however, ref. 34 found evapotranspiration was overestimated in many regions by the models. There are candidate mechanisms for issues related to runoff, for example, precipitation in models is known to lack intermittency (35, 36) which could lead to runoff being biased low if precipitation is not intermittent enough, leading to biases in water availability from the land surface, although model biases in the runoff dependency on precipitation and temperature are not systematically of one sign (37). Relatedly, ref. 38 recently documented that while the moisture transport into some arid regions from the oceans has increased, the amount of precipitation recycling in ERA5 has declined, so it is possible that models are somehow misrepresenting the role of precipitation recycling or its changes. Missing feedbacks between the land surface and the climate is another potential candidate. For example, it has recently been shown that wildfire occurrence alters runoff characteristics (39), and interactions such as these are not currently represented in climate models. Both altered transports of moisture from the ocean and altered availability of water from the land surface have probably contributed to historical trends in different regions (40), and the warming of the oceans is very likely an important control on atmospheric humidity over land (41). However, when it comes to understanding why atmospheric humidity has not risen on average and why models might not be representing this, even when observed SSTs are prescribed, we currently favor the hypotheses related to water availability from the land over atmospheric transport issues, given the global nature of the problem and the fact that it arises in regions that are influenced by widely varying dynamics and moisture sources. There is some support for this from ref. 40 which shows large-scale declines in evapotranspiration over most arid/semi-arid regions and argued that reduced water availability from the land surface has been a driver in historic trends over the La Plata basin of Argentina, in

particular. However, future work to understand this issue should continue to consider both this possibility and the potential role for issues in atmospheric transport.

While the implications of this discrepancy will not be truly known until our models can be improved, they are potentially severe. If, in reality, the rising saturation vapor pressure will continue without partial compensation from rising actual vapor pressure, then future increases in VPD and, therefore, wildfire danger (2, 5, 26) may be greater than predicted based on climate models in arid/semi-arid regions. There may also be implications for heat extremes. Given that evapotranspiration tends to act as a mediator of temperature variability (42) and, in the tropics, there are theoretical linkages between relative humidity and temperature change (13, 43), we may expect that if evapotranspiration and/or relative humidity are reduced more in reality, then heat extremes may become more severe. Indeed, over the Southwest, it has been observed that reductions in humidity are greatest on the hottest days in the historical record (44).

Conclusions

Here, we have demonstrated a major discrepancy between observation-based and climate model-based historical trends in near-surface atmospheric water vapor in arid and semi-arid regions. Climate models suggest that water vapor should have increased over arid/semi-arid regions at a rate that is close to that expected from Clausius–Clapeyron scaling, on average. In reality, atmospheric water vapor over arid and semi-arid regions has essentially remained constant over the last four decades on average, which is strongly at odds with our model-based expectations. A discrepancy between modelled and observed humidity trends is also present in more humid regions, albeit to a lesser extent, and only during the most arid times of the year. In arid/semi-arid regions, warming-driven increases in saturation vapor pressure combined with near zero trends in specific humidity are leading to increases in vapor pressure deficit which is an important driver of water stress on ecosystems and wildland fire, and models are not capturing the role of specific humidity trends in this correctly. This represents a major gap in our understanding and in climate model fidelity that must be understood and fixed as soon as possible in order to provide reliable hydroclimate projections for arid/semi-arid regions in the coming decades.

Materials and Methods

Observation-Based Datasets. We focus on the 1980 to 2020 period as it is the period that is common to all observation-based datasets.

ISD and HadISDH. We use two datasets based on the Integrated Surface Database (7) (ISD) station data. The HadISDH dataset (8) is a homogenized version of the ISD data that has been produced by making certain choices about temporal coverage and by applying homogenization procedures. In addition, we use the raw ISD station data after quality control but without homogenization to check that conclusions are not sensitive to the choices made about required temporal coverage, homogenization procedures, or how the diurnal cycle is dealt with. These are referred to as HadISDH and ISD throughout and are now described in more detail.

ISD. We use the hourly Integrated Surface Database (7) (ISD) station records of 2-m temperature and 2-m dew point temperature. Starting from 29555 available stations, we first identified stations for which records began prior to January 1980 and ended after December 2020 and for which the longitude and latitude information was available, which left 6509 stations. For each of these stations we then removed data that failed cross-checks (SOURCE flags 2, A and B), were summary observations created by NCEI that may not share the same data source flag (SOURCE flag 0), were missing (SOURCE flag 9 or REPORT_TYPE 99999) or

for which no quality control was applied (QUALITY_CONTROL flag V01). Further, air temperature and dew point temperature values with the following flags were removed: (2 = Suspect, 3 = Erroneous, 6 = Suspect from an NCEI source, 7 = Erroneous from an NCEI source, A = Data value flagged as suspect, but accepted as a good value, C = Temperature and dew point reported in whole degrees Celsius). After performing this quality control, we removed stations that had at least 1 y with no data in the 1980 to 2020 record, which left 3506 stations.

To deal with inhomogeneities in the sampling of observations over the diurnal cycle both within and across stations, we linearly interpolate the quality-controlled data points to hourly values (on the hour). We then form daily averages from these linearly interpolated values but only retain data on days for which there are at least four observations available from which to compute the interpolated diurnal cycle. We then form monthly averages from the days available in each month and define "bad months" as those for which there are more than 15 d without data (i.e., where more than approximately half of the days in the month had less than 4 valid observations). If a station has at least 48 "bad months" in the 492-mo record from January 1980 to December 2020, then it is dropped from the analysis. This left 2581 stations overall, and of those, 557 were in arid/semi-arid locations and 49 were in the US Southwest.

Each of these choices are subjective and we choose the thresholds of requiring at least 15 d with at least 4 valid observations in a month for at least 444 out of the 492 mo record to retain a sufficiently broad coverage over arid/semi-arid regions. *SI Appendix*, Fig. S1 provides an indication of the stations that remain as a function of the minimum number of observations required in a day and the number of bad months allowed in the record. Our choice of using the 4 observations per day threshold is motivated by allowing us to retain valuable stations in Southern Africa, Australia, Northern Argentina, and Northern Asia (compare *SI Appendix*, Figs. S1 *D* and *E*). However, most of the stations in the United States have zero bad months with more than 18 observations per day (*SI Appendix*, Fig. S1B), and many of the stations in Spain have less than 12 bad months in the record with a 12 observations per day threshold (*SI Appendix*, Fig. S1C). So, while we have chosen thresholds that are liberal enough to allow for a sufficient number of stations across the arid/semi-arid climate regions, many of the stations in the US Southwest and Spain, would pass much stricter thresholds.

Given data availability issues in ISD for surface pressure, in the calculation of specific humidity from ISD dew point temperatures, we use the surface pressure from the closest grid point of ERA5 on the 1/4° longitude-latitude grid, but the vapor pressure trends discussed here are entirely attributed to the water vapor content, not the surface pressure trend in ERA5 (*SI Appendix*, Fig. S3).

HadISDH. HadISDH (8) is based on HadISD (45) which is a quality-controlled dataset based on the ISD stations. The criterion used for sufficient data to form a monthly mean in HadISDH is that first the monthly mean for each hour of the day is computed only if there are at least 15 d present within the month. A monthly mean value for that station is then only computed if there are at least 4 h with a monthly mean per day with at least one in each 8-h tercile 00:00 to 08:00, 08:00 to 16:00, 16:00 to 24:00 UTC). As such, the requirements of the amount of valid data for computing a monthly mean are similar to what we use for ISD, but the diurnal cycle is dealt with in a different way. In addition, homogenization of the station time series is performed by detecting change-points and, where warranted, performing adjustments based on neighbouring stations. We use the homogenized station data as opposed to the gridded product of HadISDH so that we can treat the Southwest averaging in a similar way to ISD and follow similar criteria to that described for ISD to determine whether a station has sufficient data, but we do show the gridded product in *SI Appendix*. If a station has more than 48 missing months, then it is omitted from the analysis and if a station has a whole year of missing data in the record, it is also omitted. We only use the vapor pressure data from HadISDH as there were more gaps in the records for temperature and dew point temperature (See figure 3 of ref. 8). For vapor pressure, 2457 stations have sufficient data coverage, and of those, 559 are in arid/semi-arid locations and 86 were in the US Southwest.

Station Data Processing. Annual averages are computed by first removing the monthly seasonal cycle for each station, then averaging over the months for which data are available, and then adding back the climatological annual average. This ensures that the annual average is not impacted by unevenly sampling the seasonal cycle when months are missing. The climatological seasonal cycle is determined as the climatological average over data available for a given month

from 1980 to 2020. Annual averages are computed prior to averaging over the Southwest to ensure that each of the Southwest stations has a value for each timestep (this would not be true if spatial averaging were first performed on the monthly data). For the monthly trends averaged over the Southwest, the spatial averaging over the Southwest is performed on the monthly values and the months for which stations have missing data are assumed to have zero anomaly from the climatology. To calculate an area average of the ISD station data over the US southwest, a voronoi tesselation of stations over the contiguous United States was used to assign weights to each station within the six state region that covers California, Nevada, Utah, Colorado, New Mexico, and Arizona (red outline in Fig. 1A).

Station Data Fidelity. The majority of the ISD stations are located at airports and changes in instruments and perhaps station location have occurred over time. It is also known that inaccuracies in psychrometer measurements are highest at low humidities (8). For the US station records, ref. 29 investigated humidity trends using stations from 1961 to 1990 and performed a number of checks to assess the role of changing observing practices over that period. They concluded that such artefacts were not impacting their trends. They also noted that increased aviation traffic at the airports, where the stations are located, could lead to changes in water vapor through fuel combustion. While they found this to be a very small effect, it is worth noting that, if anything, this would lead to an increase in water vapor as opposed to a decrease. Over this period, ref. 29 were actually investigating an increase in humidity over the Southwest, but this is because they were primarily covering the period prior to 1980 over which precipitation in the region was increasing. Ref. 30 also investigated inhomogeneities in US station records over a longer period from 1930 to 2010 and noted breakpoints in the dew point temperature time series in the mid 1980s and mid 2000s associated with instrument changes, but they found this only in a small portion of the stations. They also concluded that urbanization does not appear to have biased the observations. Over China, ref. 46 found breakpoints in the time series of humidity measurements around the year 2000 when a switch from manual to automated measurements occurred, and it is interesting to note that China is one region where ERA5 lies outside of the model distribution in Fig. 3D, but the region is characterized as humid, not arid/semi-arid.

For our purposes, to investigate the potential role of changes in observing practices, sensors, or station locations, a range of running vapor pressure trends with varying start points and lengths are shown in *SI Appendix*, Fig. S7 over the US Southwest and *SI Appendix*, Fig. S21 shows the equivalent for ERA5 averaged over all arid/semi-arid regions. These indicate considerable homogeneity in the bias in trends between observations and CMIP6 over the record. For example, in the Southwest, the 30-y trend centered on 2005, which misses the mid-1980s when the HO-83 hygrothermometer model was introduced (29, 30), is significantly different between the models and observations. Similarly, a number of the trends in the lower left quadrant of the diagram which miss the mid-2000's introduction of the Vaisala relative humidity sensors (30) are significantly different from the modeled trends, although significance here is more challenging given the shorter length of the trend calculation and the impact of the 1997/1998 El Niño. Overall, given the fairly homogeneous nature of the trends over time, the lack of clear break points in the time series, the global nature of the trend discrepancy, and prior studies that have found that changes in observing practices, urbanization, or airport use have not significantly impacted on results (29, 30), we conclude that it is unlikely that artifacts within the station-based observations are dominating in the observed trends.

ERA5. We use 2-m dew point temperature and temperature, surface pressure, and specific humidity on pressure levels from ERA5 (9). Specific humidity and saturation specific humidity are calculated from 2-m dew point temperature and temperature, as outlined in the *Humidity and Vapor Pressure Calculations*. ERA5 is the latest generation reanalysis product from the European Center for Medium Range Weather Forecasts (ECMWF). Of relevance to this analysis is that ERA5 assimilates in situ station-based humidity measurements as well as humidity measurements from radiosondes, dropsondes, and aircraft as well as satellite radiance measurements that are sensitive to humidity and temperature. Scatterometer estimates of near-surface soil moisture are also assimilated into ERA5. ERA5 has been shown to exhibit discrepancies over tropical oceans prior to the mid-1990s when compared with microwave sounding estimates of

tropospheric humidity, but it compares well over tropical land (18), and globally it compares well with radio occultation measurements and other reanalyses after 1995 but global discrepancies in water vapor trends are found between different reanalysis products prior to that (11). ERA5 is available on a 0.25° longitude by 0.25° latitude grid but, unless otherwise stated, we regrid to the ∼1° CESM grid prior to analysis.

MERRA2. VP trends for the MERRA2 (47) assimilation product are shown in the *SI Appendix*. MERRA2 is NASA's latest generation reanalysis product. Of relevance to this work is that MERRA2 does not assimilate in-situ station-based humidity measurements or near-surface satellite-based radiance measurements.

JRA55. JRA55 (48), the latest reanalysis from the Japan Meteorological Agency, is also shown in *SI Appendix*. It does assimilate in-situ station-based humidity measurements, but according to ref. 48, it does so using datasets and procedures from an older generation of ECMWF reanalysis, ERA40 (49), two generations prior to ERA5 so it is possible that the assimilation of in-situ measurements has improved in ERA5 compared to this dataset.

Comparison of Reanalysis Trends. *SI Appendix*, Figs. S2 and S6 show a comparison of ISD, HadISDH, ERA5, MERRA2, and JRA55 over the US southwest and globally, respectively. ERA5 compares very well with the ISD stations. While there are reasons to be cautious about trends in any reanalyses, this close correspondence between ERA5 and the stations motivates us to trust ERA5 more than other reanalysis datasets. MERRA2 does not resemble ERA5 or ISD. The vapor pressure trends over arid/semi-arid regions in MERRA2 are relatively more positive compared to those datasets. In fact, over the Southwest, MERRA2 sits within the model distribution. It is, perhaps, unsurprising that MERRA2 more closely resembles the free running model simulations because it does not assimilate in situ humidity measurements from stations. JRA55 does assimilate in situ humidity measurements but using older datasets and procedures than ERA5. JRA55 more closely resembles ERA5 and ISD than MERRA2 over most semi-arid regions, but it does not agree well with those datasets over the US southwest, for reasons that are unclear. Overall, there are good reasons to not trust MERRA2's humidity trends and given the close agreement between ERA5 and the ISD stations, we focus our analysis on ERA5 and ISD/HadISDH.

Precipitation Datasets (GPCP, GPCC, and CRUTS). We use three observation-based precipitation datasets. Global Precipitation Climatology Project (GPCP) is a satellite and rain gauge analysis on a 2.5° longitude by 2.5° latitude grid (50). The GPCC dataset is a rain gauge-based dataset on a 1° longitude by 1° latitude grid (51). The Climatic Research Unit Timeseries (CRUTS) dataset is a rain gauge-based dataset on a 0.5° longitude by 0.5° latitude grid (52).

Berkeley Earth Surface Temperature (BEST). To provide an additional estimate for Saturation Vapor Pressure trends given discrepancies in near-surface air temperature over the US Southwest between ERA5 and ISD (Fig. 1C), we also use the BEST near-surface air temperature dataset (53) on a 1° longitude by 1° latitude grid. This agrees more with ISD than with ERA5.

CMIP6. We use the Historical and Shared Socioeconomic Pathway 5-8.5 (SSP5-8.5) simulations from the Coupled Model Intercomparison Project phase 6 (CMIP6) (54). These simulations begin in 1850 and are run under historical forcings prior to 2015 and SSP5-8.5 forcings, thereafter. We use monthly mean near-surface air temperature (CMIP6 variable *tas*), near-surface specific humidity (*huss*), specific humidity on pressure levels (*hus*), surface pressure (*ps*), and precipitation (*pr*). In total, 201 members from 41 models have the necessary data available from the historical and SSP5-8.5 simulations as summarized in *SI Appendix*, Table S1.

AMIP6. The Atmospheric Model Intercomparison Project (AMIP) simulations from CMIP6 (AMIP6) are run with historical forcing and prescribed observation-based SSTs. These simulations end in 2014, so we only use the period 1980 to 2014 and do not include these simulations for trend calculations in the main text where we focus on 1980 to 2020 but we show 1980 to 2014 trends in the

SI Appendix and the time series from these simulations in Fig. 5. 117 members from 35 different models are used (*SI Appendix*, Table S2).

CESM2. The Community Earth System Model 2 (CESM2) (55) is a CMIP6-era model, and here we use the CESM2 large ensemble (56) (LENS2), which is a 100 member coupled ensemble that is run under CMIP6 forcings from 1850 to 2100 using historical forcings prior to 2015 and SSP3-7.0 forcings thereafter. We also use a 10-member ensemble of CESM2 Global Ocean Global Atmosphere (GOGA2) simulations (57), in which observation-based SSTs from ERSSTv5 (58, 59) are prescribed alongside the same forcings as used in LENS2 (historical prior to 2015 and SSP3-7.0 thereafter).

CMIP5. CMIP5 models are shown in *SI Appendix*, Fig. S20 using the historical simulations prior to 2006 and the RCP8.5 scenario thereafter for the models and members summarized in *SI Appendix*, Table S3.

Data Pre-Processing. Unless otherwise stated, all gridded observation and model-based datasets were regridded to the ∼1° horizontal grid of CESM2 using bilinear interpolation prior to any analysis. For the CMIP6 data, whenever an ensemble mean is shown, the average over the ensemble members available for each model is first calculated followed by the mean across models, such that each model is given equal weight, albeit with greater noise contributions from models with fewer members. However, when comparing the observations to the model distribution, the spread across model members from all models is used. While this means that models that have more ensemble members available are given more weight in the analysis, this maximizes the chance of the model spread encompassing the observations. For all datasets, monthly means were first derived before any vapor pressure or humidity calculations were performed. See below for the form of these calculations and for a demonstration that results are insensitive to this choice. Humidity-related quantities and precipitation were then averaged into annual means and, for trends, the ordinary least squares trend was computed from 1980 to 2020 on these annual means.

Humidity and Vapor Pressure Calculations. We derive the various quantities related to atmospheric water vapor following the relationships provided by ref. 60. For ISD, ERA5, and JRA55, the vapor pressure (VP), in hPa, is calculated from 2-m dew point temperature (T_D), in °C, using:

$$VP = 6.112 \times \exp\left(\frac{17.67 T_D}{T_D + 243.5}\right).$$ [1]

Specific humidity (q) can then be calculated from VP using:

$$q = \frac{0.622 VP}{ps - 0.378 VP},$$ [2]

where ps is surface pressure, in hPa.

Similarly, for all datasets, the saturation vapor pressure is calculated by replacing T_D in Eq. **1** with near-surface air temperature and the saturation specific humidity is calculated by replacing VP in Eq. **2** with the saturation vapor pressure.

For the model simulations and MERRA2, near-surface (2-m) specific humidity is the variable that is provided, so the vapor pressure is calculated by rearranging Eq. **2** to give:

$$VP = \frac{q \times ps}{0.622 + 0.378 q}.$$ [3]

For JRA55 for which both T_D and 2 m q are available, we checked that it did not matter whether VP was computed from T_D or from q (*SI Appendix*, Fig. S2) and our further use of column specific humidity from ERA5 in Fig. 1F provides confidence that potential inaccuracies in the above empirical relationships are not playing a role. Furthermore, while the calculation of VP also includes surface pressure, it is the humidity trends that are leading to trends in VP, not trends in surface pressure (*SI Appendix*, Fig. S3).

These humidity-related quantities were derived using monthly mean variables. We checked for ERA5 that the non-linearities present in the above equations did not lead to different conclusions if vapor pressure was calculated

from hourly or daily averaged dew point temperature prior to computing the monthly average (*SI Appendix*, Fig. S4) and, aligned with the results of ref. 27, we found that the difference was minimal. We also show in *SI Appendix*, Fig. S4 that for ERA5 the same conclusions are drawn if the vapor pressure calculation is performed using dew point temperature on the native grid.

Incorporation of Precipitation Trends. Denoting the 1980 to 2020 trends with $\overrightarrow{(.)}$, we perform an ordinary least squares linear regression across model members (i), i.e.,

$$\overrightarrow{VP}(i) = a + b \times \overrightarrow{pr}(i) + \epsilon(i),$$ [4]

where a and b are the regression coefficients, pr is precipitation, and $\epsilon(i)$ represents the residuals from the regression fit. The regression slope (b) at each grid point can be seen in *SI Appendix*, Fig. S10*D*, and this is also similar to the slope of the relationship between observed interannual variability in VP and pr (*SI Appendix*, Fig. S10*E*). We quantify the uncertainty around the CMIP6 regression fit using the $\pm 2\sigma(\epsilon)$ range, where σ refers to the SD. A demonstration of this method is given in Fig. 2*A* for the Southwest, where the black line indicates the regression fit and the $\pm 2\sigma(\epsilon)$ range is given by the gray shading. In Fig. 3*D*, we take this approach at every land grid point and plot the difference between the ERA5 VP trend and that which would be predicted based on Eq. **4** using precipitation from GPCC. A similar analysis is shown for the other observed precipitation datasets in *SI Appendix*, Fig. S11.

Aridity Index and Defining Aridity Zones. We calculate the Aridity Index (AI) as the ratio of climatological (1980 to 2020, annual mean) precipitation to potential evapotranspiration using precipitation and potential evapotranspiration from the TerraClim dataset (61) and depict the following aridity zones in Fig. 3*A*: Hyper-arid (AI < 0.05); Arid/Semi-Arid ($0.05 \leq$ AI < 0.5); Humid (AI $\geq$ 0.5). In Figs. 4 and 5, only the grid points that are defined as Arid/Semi-Arid by this definition after interpolation onto CESM2's $\sim$1° grid are shown, and in Fig. 3 *E–M*, this annual mean AI is used to bin spatial locations into aridity bins. In Fig. 3 *K–M*, a monthly AI is used in the ordering of months from most arid to most humid. This monthly AI is calculated as the ratio of the 1980 to 2020 averaged precipitation to potential evapotranspiration but for each month of the year separately. In some months of the year, potential evapotranspiration becomes zero (or close to zero), leading to infinite or very large values of Aridity Index, so for this reason, we cap the monthly AI at values of 10.

Binning of Trends in Fig. 3. To produce panels Fig. 3 *E–J*, we take data on all land surface grid points excluding Antarctica on the $\sim$1° CESM2 grid and divide them into 30 equal-area bins based on the observed annual mean Aridity Index, described above. For each Aridity Index bin, we then sort the grid points into 15 equal-area bins based on their precipitation trends (using the precipitation from each model, and GPCC for the observation-based precipitation). The result is that the same locations are not present in each bin for each dataset because their precipitation trends differ. We then compute the area weighted average of the grid points in these bins. To normalize the specific humidity trend by temperature trend, we divide by the area weighted average temperature trend over land (excluding Antarctica). *SI Appendix*, Fig. S13 shows replicas of Fig. 3 *H–J* but using GPCP and CRUTS precipitation. *SI Appendix*, Fig. S13 also demonstrates that the normalized results are similar when instead normalizing by the local temperature trend, but differ when normalizing by the global mean temperature including oceans. *SI Appendix*, Fig. S14 demonstrates that the precipitation and temperature trends in each bin in observations lie within the CMIP6 model distribution.

To produce Fig. 3 *K–M*, we use trends from 1980 to 2020 for each month of the year and for each land surface grid point on the $\sim$1° CESM2 grid. At each

location, we order the months of the year according to the observed monthly Aridity Index and then we bin all the land spatial grid point into 30 equal-area bins according to the annual mean Aridity Index.

Data, Materials, and Software Availability. All datasets used in this study are freely available at the following websites: ISD: https://www.ncei.noaa.gov/data/global-hourly/archive/csv/ (7); ERA5: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 (9); MERRA2: https://disc.gsfc.nasa.gov/datasets?project=MERRA-2 (47); JRA55: http://jra.kishou.go.jp/JRA-55/index_en.html (48); BEST: https://berkeleyearth.org/data/ (53); CMIP6: https://esgf-node.llnl.gov/search/cmip6/ (54); LENS2: https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html (57); GOGA2: https://www.cesm.ucar.edu/working-groups/climate/simulations/cam6-prescribed-sst; GPCC: https://psl.noaa.gov/data/gridded/data.gpcc.html (51); GPCP: https://psl.noaa.gov/data/gridded/data.gpcp.html (50); and CRUTS: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.05/ (52). All datasets used in this study are freely available as described in each reference of the methods section. The analysis codes to reproduce the figures of this manuscript are available at https://github.com/islasimpson/qtrend_paper (62), and the data required to reproduce the figures is available at https://doi.org/10.5065/5jz0-0n57 (63).

Author affiliations: aClimate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80305; bDepartment of Statistics and Data Science, University of California, Los Angeles, CA 90095; cDepartment of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095; dInstitute of the Environment and Sustainability, University of California, Los Angeles, CA 90095; eEarth Research Institute, University of California, Santa Barbara, CA 93106; fDepartment of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY 14850; gPolar Bears International, Bozeman, MT 59772; and hOcean and Climate Physics, Lamont-Doherty Earth Observatory, Columbia University, New York, NY 10964

1. J. S. Mankin *et al.*, "NOAA drought task force report on the 2020–2021 southwestern U.S. drought" (Tech. Rep., NOAA Drought Task Force, MAPP, and NIDIS, 2021).
2. P. R. Higuera, J. T. Abatzoglou, Record-setting climate enabled the extraordinary 2020 fire season in the western United States. *Glob. Change Biol.* **27**, 2021 (2020).

3. A. P. Williams *et al.*, Large contribution from anthropogenic warming to an emerging North American megadrought. *Science* **368**, 314–318 (2020).
4. A. P. Williams, B. I. Cook, J. E. Smerdon, Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nat. Clim. Change* **12**, 232–234 (2022).

5. J. T. Abatzoglou, A. P. Williams, Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Ac. Sci. U.S.A.* **113**, 11770–11775 (2016).

6. T. W. Jacobson, R. Seager, A. P. Williams, N. Henderson, Climate dynamics preceding summer forest fires in California and the extreme case of 2018. *J. Appl. Meteorol. Climatol.* **61**, 989–1002 (2022).

7. A. Smith, N. Lott, R. Vose, The integrated surface database: Recent developments and partnerships. *Bull. Amer. Met. Soc.* **92**, 704–708 (2011).

8. K. M. Willett *et al.*, HadISDH land surface multi-variable humidity and temperature record for climate monitoring. *Clim. Past* **10**, 1983–2006 (2014).

9. H. Hersbach *et al.*, The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

10. I. M. Held, B. J. Soden, Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).

11. H. Douville, S. Qasmi, A. Ribes, O. Bock, Global warming at near-constant tropospheric relative humidity is supported by observations. *Nat. Commun. Earth Environ.* **3**, 1–7 (2022).

12. M. M. Joshi, J. M. Gregory, M. J. Webb, D. M. H. Sexton, T. C. Johns, Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. *Clim. Dyn.* **30**, 455–465 (2008).

13. M. P. Byrne, P. A. O'Gorman, Link between land-ocean warming contrast and surface relative humidities in simulations with coupled climate models. *Geophys. Res. Lett.* **40**, 5223–5227 (2013).

14. S. Sherwood, Q. Fu, A drier future? *Science* **343**, 737–739 (2014).

15. P. A. O'Gorman, C. J. Muller, How closely do changes in surface and column water vapor follow Clausius-Clapeyron scaling in climate change simulations? *Env. Res. Lett.* **5**, 025207 (2010).

16. A. J. Simmons, K. M. Willett, P. D. Jones, P. W. Thorne, D. P. Dee, Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.* **115**, D011110 (2010).

17. R. J. H. Dunn, K. M. Willett, A. Ciavarella, P. A. Stott, Comparison of land surface humidity between observations and CMIP5 models. *Earth Syst. Dyn.* **8**, 719–747 (2017).

18. R. P. Allan, K. M. Willett, V. O. John, T. Trend, Global changes in water vapor 1979–2020. *JGR Atmos.* **127**, e2022JD036728 (2022).

19. S. I. Seneviratne, M. G. Donat, B. Mueller, L. V. Alexander, No pause in the increase of hot temperature extremes. *Nature Clim. Change* **4**, 161–163 (2014).

20. H. Douville, M. Plazzotta, Midlatitude summer drying: An underestimated threat in CMIP5 models. *Geophys. Res. Lett.* **44**, 9967–9975 (2017).

21. R. N. Jones, J. H. Ricketts, Regime changes in atmospheric moisture under climate change. *Atmosphere* **13**, 1577 (2022).

22. H. Douville, K. M. Willett, A drier than expected future, supported by near-surface relative humidity observations. *Sci. Adv.* **9**, eade6253 (2023).

23. J. Scheff, J. C. Burroughs, Diverging trends in US summer dewpoint since 1948. *Int. J. Climatol.* **43**, 4183–4195 (2023).

24. F. Lehner, C. Deser, I. R. Simpson, L. Terray, Attributing the U.S. Southwest's recent shift into drier conditions. *Geophys. Res. Lett.* **45**, 6251–6261 (2018).

25. R. Seager *et al.*, Mechanisms of a meteorological drought onset: Summer 2020 to spring 2021 in Southwestern North America. *J. Clim.* **35**, 3767–3785 (2022).

26. F. Sedano, J. T. Randerson, Multi-scale influence of vapor pressure deficit on fire ignition and spread in boreal forest ecosystems. *Biogeosciences* **11**, 3739–3755 (2014).

27. R. Seager *et al.*, Climatology, variability, and trends in the U.S. vapor pressure deficit, an important fire-related meteorological quantity. *J. Appl. Meteorol. Climatol.* **54**, 1121–1141 (2015).

28. A. M. Chiodi, B. E. Potter, N. K. Larkin, Multi-decadal change in western US nighttime vapor pressure deficit. *Geophys. Res. Lett.* **48**, e2021GL092830 (2021).

29. D. J. Gaffen, R. J. Ross, Climatology and trends of U.S. surface humidity and temperature. *J. Clim.* **12**, 811–828 (1999).

30. P. J. Brown, A. T. DeGaetano, Trends in U.S. surface humidity 1930–2010. *J. Appl. Meteorol. Climatol.* **52**, 147–163 (2013).

31. H. Douville *et al.*, Drivers of the enhanced decline of land near-surface relative humidity to abrupt 4xCO2 in CNRM-CM6-1. *Clim. Dyn.* **55**, 1313–1629 (2020).

32. A. M. Ukkola *et al.*, Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Env. Res. Lett.* **11**, 104012 (2016).

33. M. Zhao, A. Geruo, Y. Liu, A. G. Konings, Evapotranspiration frequently increases during droughts. *Nat. Clim. Change* **12**, 1024–1030 (2022).

34. B. Mueller, S. I. Seneviratne, Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.* **14**, 128–134 (2013).

35. G. L. Stephens *et al.*, Dreary state of precipitation in global models. *JGR Atmos.* **115**, D24211 (2010).

36. K. E. Trenberth, Y. Zhang, M. Gehne, Intermittency in precipitation: Duration, frequency, intensity, and amounts using hourly data. *J. Hydrometeorol.* **18**, 1393–1412 (2017).

37. F. Lehner *et al.*, The potential to reduce uncertainty in regional runoff projections from climate models. *Nat. Clim. Change* **9**, 926–933 (2019).

38. C. Wang, J. Li, F. Zhang, K. Yang, Changes in the moisture contribution over global arid regions. *Clim. Dyn.* **61**, 543–557 (2023).

39. A. P. Williams *et al.*, Growing impact of wildfire on western US water supply. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2114069119 (2022).

40. S. M. Vincente-Serrano *et al.*, Recent changes of relative humidity: Regional connections with land and ocean processes. *Earth Syst. Dyn.* **9**, 915–937 (2018).

41. M. P. Byrne, P. A. O'Gorman, Trends in continental temperature and humidity directly linked to ocean warming. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4863–4868 (2018).

42. S. I. Seneviratne *et al.*, Investigating soil moisture-climate interactions in a changing climate: A review. *Earth Sci. Rev.* **99**, 125–161 (2010).

43. M. P. Byrne, Amplified warming of extreme temperatures over tropical land. *Nat. Geosci.* **14**, 837–841 (2021).

44. K. A. McKinnon, A. Poppick, I. R. Simpson, Hot extremes have become drier in the United States Southwest. *Nature. Clim. Change* **11**, 598–604 (2021).

45. R. J. H. Dunn *et al.*, HadISD: A quality controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Clim. Past* **8**, 1649–1679 (2012).

46. Z. Li, Z. Yan, Y. Zhu, N. Freychet, S. Tett, Homogenized daily relative humidity series in China during 1960–2017. *Adv. Atmos. Sci.* **37**, 318–327 (2020).

47. R. Gelaro *et al.*, The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).

48. S. Kobayashi *et al.*, The JRA-55 reanalysis: General specifications and basic characteristics. *J. Met. Soc. Japan.* **93**, 5–48 (2015).

49. S. M. Uppala *et al.*, The era-40 reanalysis. *Quart. J. Roy. Meteor. Soc.* **131**, 2961–3012 (2005).

50. R. Adler *et al.*, "Global precipitation climatology project (GPCP) climate data record (CDR), version 1.3 (Daily)" (Tech. Rep., 2017).

51. K. Schamm *et al.*, "GPCC full data daily version 1.0 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data" (Tech. Rep., 2015).

52. I. Harris, P. D. Jones, T. J. Osborn, D. H. Lister, Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).

53. R. A. Rohde, Z. Hausfather, The berkeley earth land/ocean temperature record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).

54. V. Eyring *et al.*, Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Mod. Dev.* **9**, 1937–1958 (2016).

55. G. Danabasoglu *et al.*, The community earth system model 2 (CESM2). *J. Adv. Mod. Earth Sys.* **12**, e2019MS001916 (2019).

56. K. B. Rodgers *et al.*, Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.* (2021).

57. CESM climate variability and change working group. CESM webpage containing information on how to download the CESM2 GOGA simulations. CESM2 GOGA simulations. https://www.cesm.ucar.edu/working-groups/climate/simulations/cam6-prescribed-sst/. Deposited 31 August 2021.

58. B. Huang *et al.*, Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).

59. B. Huang *et al.*, NOAA Extended Reconstructed Sea Surface Temperature (ERSST). Version 5 (2017).

60. D. Bolton, The computation of equivalent potential temperature. *Mon. Weather Rev.* **108**, 1046–1053 (1980).

61. J. T. Abatzoglou, S. Z. Dobrowski, S. A. Parks, K. C. Hegewisch, Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Nat. Sci. Data* **5**, 170191 (2015).

62. I. R. Simpson, Data repository containing the data required to reproduce the figures in this manuscript. Github. https://github.com/islasimpson/qtrend_paper. Deposited 19 October 2023.

63. I. R. Simpson, Observed humidity trends in dry regions contradict climate models, Version 1.0. UCAR/NCAR - GDEX. https://doi.org/10.5065/5jz0-0n57. Deposited 27 September 2023.