Data Adoption Justification Memo (for California's Fifth Climate Change Assessment)

Climate Adjusted Synthetic Relative Humidity Data

LeRoy Westerling[1], Jonathan Baldwin[1], Shane Romsos[2], and David Saah[2,3]

This project was funded by the Electric Program Investment Charge (EPIC) program, administered by the California Energy Commission: Agreement EPC-18-026 *Comprehensive open-source development of next generation wildfire models for grid resiliency* (Pyregence Project).

April 2026

Note: The synthetic relative humidity data were generated specifically to address the needs of the Pyregence Consortium for modeling long-term wildfire scenarios that the original downscaled data did not. Others may find these synthetic data will be useful for their purposes as well. However, the authors make no claims that the synthetic version is usable for other purposes. Users are strongly encouraged to make such determinations themselves.

This document provides justification for modifications to the data inputs used in the Pyregence long-term wildfire projection models, specifically regarding the treatment of relative humidity (RH). Accurate climate inputs are essential for simulating future wildfire activity and vegetation dynamics, and relative humidity is a critical variable for driving wildfire projections. However, **a detailed comparison between observed historical RH data and downscaled, bias-corrected projections from global climate models has revealed significant discrepancies in both mean values and long-term trends.** These differences undermine the physical consistency between historical and projected datasets, resulting in unrealistic fire activity outputs when the models are run using projected RH values. **This memo discusses the rationale for adjusting the modeling approach used in the Pyregence project,** including the calculations used for "synthetic" RH from climate simulation input variables, and describes the implications for use in projecting future wildfire hazards across California. This synthetic RH product was developed to remediate an issue for the Pyregence wildfire modeling but may not be appropriate for other applications. Potential users are encouraged to investigate to satisfy themselves which dataset is more appropriate for their application.

Background and Motivation

Pyregence's long-term wildfire projection team (Workgroup 4) fit preliminary models for various fire activity metrics (presence, size, severity, spread, etc.), as well as processes characterizing changes to and transfer of biomass amongst various carbon pools, by first using historical observations of climate that were publicly available, in anticipation of having official downscaled and bias-corrected historical climate data products from the UCLA and Scripps Institution of Oceanography teams (EPC-20-006). Once the downscaled, bias-corrected historical climate data from UCLA and Scripps were available, the Pyregence team reprocessed these data and then recalibrated all long-term projection models (e.g., UC Merced/Westerling statistical fire model, LUCAS, LANDIS II) to this new historical reference data set.

---

[1] UC Merced

[2] Spatial Informatics Group, LLC

[3] University of San Francisco

One required and highly anticipated data product from these CEC-funded climate data sets was downscaled and bias-corrected relative humidity climate variables. Daily minimum and maximum relative humidity, and derivative variables such as vapor pressure deficit and fuel moisture that depend on relative humidity, are extremely relevant as model inputs for characterizing fire activity. This is because wildfire activity is strongly correlated with relative humidity and is readily interpretable. Consequently, all the fire components of our modeling framework as they have been specified, whether statistical fire models or fire modules within dynamic vegetation models like LANDIS II, ultimately rely on downscaled, bias-corrected relative humidity data.

Projecting fire activity and vegetation changes into the future essentially involves taking our models (UC Merced/Westerling and LANDIS II) that have been calibrated against historical observation data products, and applying them to downscaled, bias-corrected versions of the projected climate data to simulate future fire activity and vegetation conditions. This process assumes that relative humidity derived from historical observations and relative humidity simulated by global climate models (GCMs) (once downscaled and bias-corrected) describe the same climate physics and have highly similar statistical properties when they represent systems driven by the same climatic drivers (e.g., cumulative greenhouse gas emissions).

After further exploration of projected relative humidity values simulated by GCMs, this assumption does not appear to be correct when comparing the observed historical and projected climate data sets and the simulation results that they generate for fire activity and vegetation. Relative humidity does not behave the same in the projected climate data sets as it occurred in the historical climate data sets. For the period in which these two climate data sets overlap (2000-2020), relative humidities simulated by the global climate models are significantly wetter than observed relative humidities. This difference does not appear to be rectifiable by bias-correction, because it is not just average values which are different: observed historical relative humidity in California has been characterized by a declining trend, in which ever-dryer conditions have been driving rapid increases in fire activity, which is not reflected in the simulated relative humidity for either 2000-2020 or beyond. The two data sets do not describe the same physics for regional climate (Figures 1 and 2). This situation profoundly affected our simulations of wildfire activity across the State, because it means that fire activity simulated on the bases of the simulated climate will be low and likely decreasing, whereas observed fire activity in recent decades is much higher and increasing. This means that, regardless of the global emissions scenario, climate model, fuels management scenario, or development footprint scenario, projected fires will decrease immediately and dramatically.

One consequence of a warming climate is that evaporation increases, and averaged at a global scale, the atmosphere holds more moisture over time. It appears that GCMs make assumptions about the transport of atmospheric moisture from over the oceans and more humid land areas into arid regions that have not been borne out by observations in recent decades: where arid regions do see more wet extremes, overall they have been becoming more arid with warming in recent decades, particularly in the dry season (i.e., “fire season”). Consequently, the discrepancy between simulated and observed relative humidity has been significant and increasing across all arid regions of the globe, not just California (Simpson et al., 2024).

There was no simple statistical “fix” whereby our colleagues at UCLA and/or Scripps could adjust the projected relative humidity data to more accurately reflect what we believe is correct for local conditions in California, given that these data are already bias corrected. Bias correction can adjust for modest local discrepancies in mean or variance, but not for differences in long term trends. One possible solution would have been for the Pyregence team to recalibrate ***all*** of our models to use only temperature and precipitation (and windspeed where relevant), and to exclude relative humidity from all of our models. This would have been more easily done in some models than others. LANDIS II, for example, requires relative humidity.

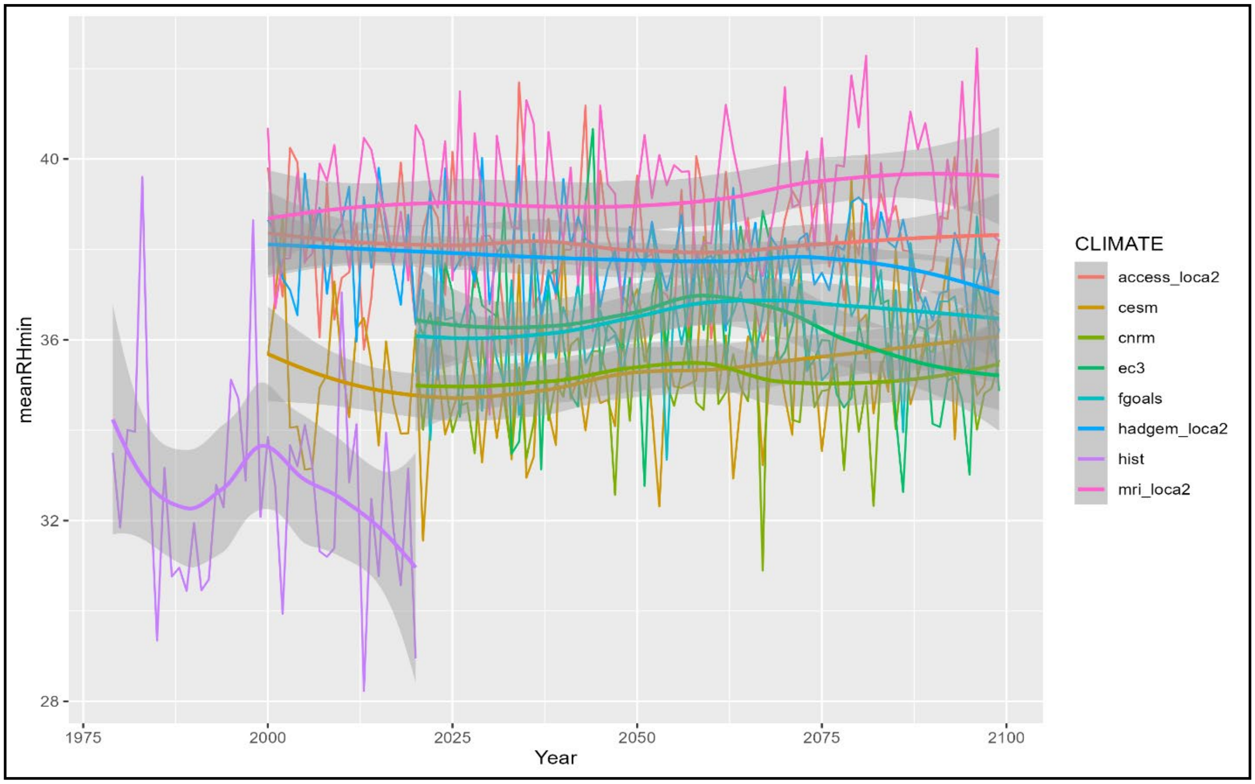

Figure 1. An example of relative humidity trends for southern Sierra Nevada region of California. The heavy purple line shows the mean relative humidity as modeled from historical ERA 5 WRF data from 1980-2020, showing a declining trend. The other lines represent the results from the LOCA2-Hybrid downscaled climate projections from Scripps for various global climate models from 2000-2100. For the period in which these two climate data sets overlap (2000-2020), relative humidities simulated by the global climate models are significantly wetter than observed relative humidities and do not reflect the observed downward trend.

Instead, we chose to estimate statistical regression models for daily minimum and maximum relative humidity as functions of daily precipitation and daily minimum and maximum temperature. The rationale for this is simple - relative humidity is a function of the moisture "inputs" and "outputs" in any given location, and moisture lost from evapotranspiration (which depends on temperature) and net moisture gained from precipitation and transport. The statistical modeling approach we used does not capture all the effects on relative humidity in each location from transport of atmospheric moisture, beyond the extent to which they are correlated with locally observed daily precipitation. Consequently, similar to relative humidity estimated for previous California state climate assessments with the Variable Infiltration Capacity (VIC) hydrologic model, we do not capture part of the variation in daily humidity driven by large scale atmospheric moisture transport (Pierce and Westerling, 2013). The impact of this limitation is much greater for wet extremes of daily maximum relative humidity than it is for daily minimum relative humidity, which consequently mutes the impact on the statistical fire models. This is because these models generally are not predicting any fire at daily relative humidities far below the wet extremes that we underestimate. That is, our approach underestimates wet extremes of daily maximum humidity, but since fire models are most sensitive to minimum humidity and generally do not predict fire activity at high humidity levels, this limitation has limited impact on predicting fire activity. We also do not have a representation of daily dew point temperature, and consequently our maximum relative humidity simulations do not capture spikes to 100% relative humidity when dew is increasing. Despite these limitations, we are able to capture a significant portion of the variability in daily relative humidity, as well as the trends observed in recent decades, and because the fire models are most sensitive to daily minimum relative humidity, we believe this is an acceptable solution. Notably, fire models that exclude relative humidity and are recalibrated to rely only on temperature, precipitation, and wind face the same limitations.

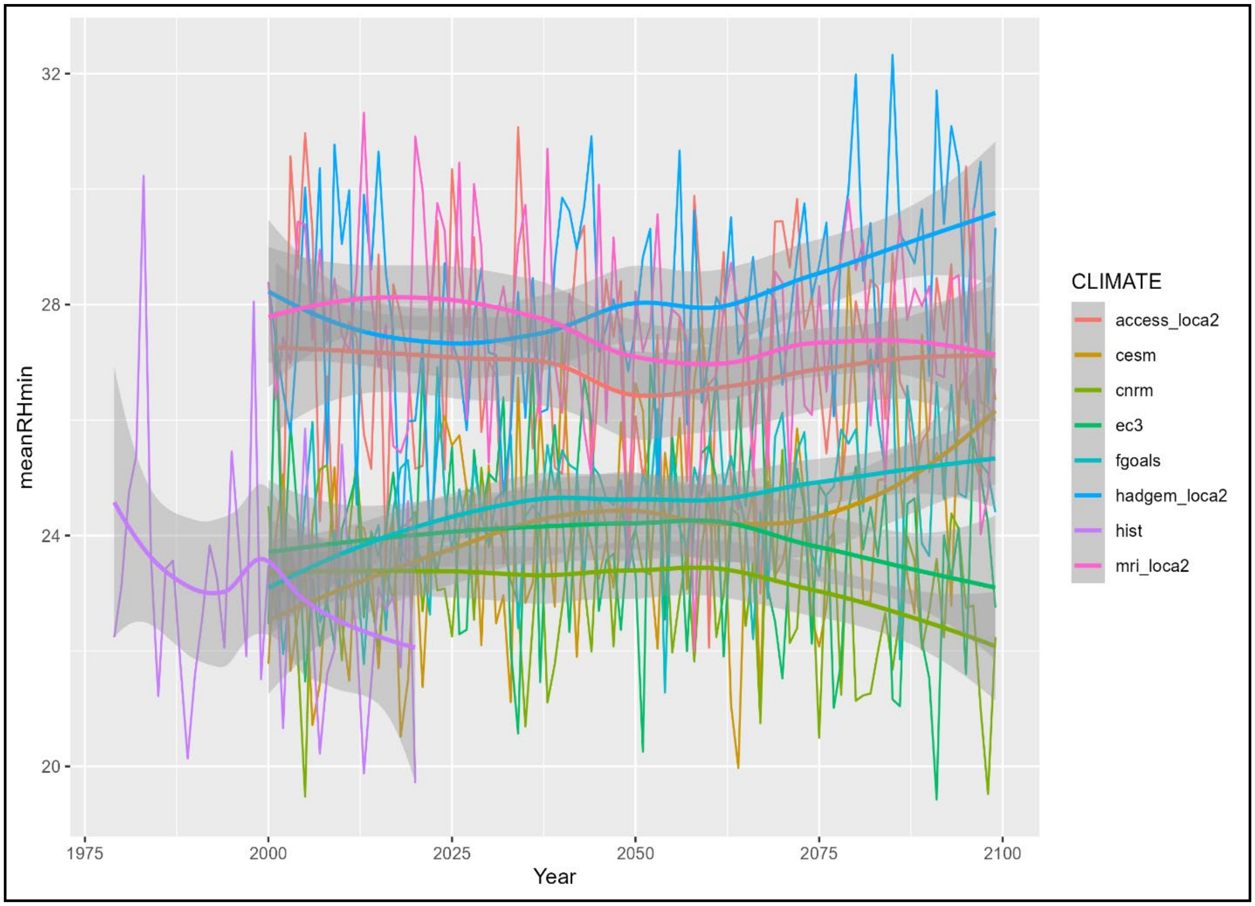

Figure 2. An example of relative humidity trends for the Klamath region of California. The heavy purple line shows the mean relative humidity as modeled from historical ERA 5 WRF data from 1980-2020, showing a declining trend. The other lines represent the results from the LOCA2-Hybrid downscaled climate projections from Scripps for various global climate models from 2000-2100. For the period in which these two climate data sets overlap (2000-2020), relative humidities simulated by the global climate models are significantly wetter than observed relative humidities and do not reflect the observed downward trend.

Synthetic Relative Humidity Model Estimation Methods

To address the limitations of existing downscaled and bias-corrected projected relative humidity (RH) data for fire modeling applications, we developed a robust statistical modeling approach to estimate daily minimum and maximum RH values using historical climate observations and landscape characteristics. Historical climate observations used to fit the statistical model were derived from downscaled and bias corrected climate data (CMIP6 GCMs downscaled using WRF) sourced from the UCLA Center for Climate Science (Rahimi, 2022, Rahimi et al., 2024). These data are accessible at the *Registry for Open Data on AWS* (https://registry.opendata.aws/wrf-cmip6/) and are managed by UCLA Center for Climate Science.

Working in the R statistical programming environment, we tested semiparametric models using the bam() function in the mgcv library with smoothed interactive terms as explanatory variables for daily minimum and maximum relative humidity. This approach allows for nonlinear fits to accommodate complex interactions between variables and has been used previously for applications like fire modeling (e.g., Preisler and Westerling, 2007, Westerling and Bryant, 2008, Preisler et al., 2011, Westerling et al., 2011a, Westerling et al., 2011b, Bryant and Westerling, 2014). Because the number of historical daily observations is very large (more than 1 billion), we worked with a random subsample of 1 million observations to estimate our models. This is a more than adequate sample size to robustly capture the variability in relative humidity in California and its drivers.

Daily minimum and maximum relative humidity are expressed as fractions between 0 and 1. To allow for more efficient modeling using our chosen modeling framework, we transformed these fractions to more normally distributed continuous unbounded variables using the formulas:

$$\mathrm{rmin2} = \mathbf{log}((\mathrm{rmin})/(1-\mathrm{rmin}))$$

$$\mathrm{rmax2} = \mathbf{log}((\mathrm{rmax})/(1-\mathrm{rmax}))$$

where rmin and rmax are daily relative minimum and maximum relative humidity expressed as fractions, and rmin2 and rmax2 are their transformations.

We tested a restricted suite of explanatory variables, given our understanding of relative humidity and given what variables would be available from downscaled climate projections:

tmx, tmn, prc:	gridded 3 km daily maximum and minimum temperature and daily cumulative total precipitation,
wsp	gridded 3 km daily average windspeed,
t30, p30:	daily mean temperature and cumulative total precipitation averaged over the preceding 30 days,
t60, p60:	daily mean temperature and cumulative total precipitation averaged over the preceding 60 days,
t90, p90:	daily mean temperature and cumulative total precipitation averaged over the preceding 90 days.

In addition to these daily climate variables, we also included fixed landscape variables:

x, y	gridded location indices at 3 km resolution within California,
elv, asp	3 km gridded average elevation and aspect.

We added variables sequentially and compared Akaieke Information Criterion (AIC) scores to choose between models with different specifications. 30, 60, and 90-day averages of precipitation and temperature were very highly correlated, so we did not consider models that included more than one duration of leading average values for precipitation or temperature.

The best fit models that we selected for daily relative humidities took the form:

```
rmin2 ~ s(tmx,tmn,prc)+s(t90,p90)+s(elv,asp)+s(x,y)
```

and

```
rmax2 ~ s(tmx,tmn,prc)+s(t90,p90)+s(elv,asp)+s(x,y)
```

where functions s(*) are spline-based, smoothed, nonlinear interactions of the included variables. This functional form estimates daily minimum or maximum relative humidity as a nonlinear semiparametric function of a smoothed interaction between daily minimum temperature, maximum temperature and precipitation; a smoothed interaction between temperature and precipitation averaged over the preceding 90 days; a smoothed interaction between elevation and aspect; and smoothed location.

We could have achieved marginal improvements in the models as indicated by their AIC scores by including an additional term for a smoothed day of the year - essentially adding a seasonal cycle to the explanatory variables - but chose not to do so, preferring that any seasonal cycle in the estimates emerge from the climate data.

The resulting models explain about 66% and 45%, respectively, of the variation in daily minimum and maximum relative humidity, and this result was very robust across randomly selected data samples.

We note that the issues we describe above with the original projected relative humidity data are likely to affect assessment work in other sectors as well that rely on relative humidity. Researchers should carefully assess whether to use our statistically derived relative humidity estimates as replacements based on the characteristics of the impacts they are modeling.

Data Access

Daily Synthetic Relative Humidity raster datasets provide spatially explicit, statewide estimates of maximum and minimum synthetic relative humidity in a raster format, enabling consistent analysis across historical and future climate scenarios. These datasets are available from multiple modeling sources - including historical ERA 5 WRF (1950–2021, used to fit statistical model), UCLA WRF Bias-Corrected projections, and LOCA2-Hybrid projections - all delivered as Cloud-Optimized GeoTIFFs using the original source data's projection and grid. Data are organized for efficient access, with historical ERA 5 WRF available directly by date, and UCLA WRF Bias-Corrected and LOCA2-Hybrid datasets structured by climate scenario (GCM/SSP/run ID), year, and date-specific files (YYYY_MM_DD). Together, these resources can support climate and wildfire research, hydrological modeling, and other environmental applications by providing daily humidity estimates across a range of temporal and scenario contexts. The following provides additional details on the datasets.

1. *Daily Synthetic Relative Humidity Rasters (ERA 5 WRF)* – This dataset is historical ERA 5 WRF data (1950 to 2021), provided as a daily, statewide raster (Cloud-Optimized GeoTiff) using the same projection and grid as the ERA 5 WRF data source. Data can be accessed from the following link - http://ungoliant.ucmerced.edu/data/Climate-Data/era5wrf/.

2. *Daily Synthetic Relative Humidity Rasters (UCLA WRF_Bias Corrected)* - This dataset is first organized by climate scenario. The data is a daily, statewide raster (Cloud-Optimized GeoTiff) using the same projection and grid as UCLA WRF Bias Corrected source data. Data can be accessed from the following link - http://ungoliant.ucmerced.edu/data/Climate-Data/wrf_bc/. The directory is structured by gcm/ssp/runid name set and then by Daily/<year>/synrh<max/min>_<date>.tif. The files are synrh (synthetic relative humidity) followed by max or min. The filename is organized by YYYY_MM_DD.
3. *Daily Synthetic Relative Humidity Rasters (LOCA2-Hybrid)* - This dataset is first organized by climate scenario. The data is a daily, statewide raster (Cloud-Optimized GeoTiff) using the same projection and grid as the LOCA2-Hybrid source data. Data can be accessed from the following link - http://ungoliant.ucmerced.edu/data/Climate-Data/loca2/ . The directory is structured by gcm/ssp/runid name set and then by Daily/<year>/synrh<max/min>_<date>.tif. The files are synrh (synthetic relative humidity) followed by max or min. The filename is organized by YYYY_MM_DD.

Conclusion

In conclusion, the development of synthetic relative humidity datasets addresses critical limitations in downscaled, bias-corrected climate projections that significantly misrepresent historical RH trends in California, leading to unrealistic wildfire model outputs. By statistically estimating daily minimum and maximum RH from temperature, precipitation, and other environmental variables, the Pyregence team has created a physically consistent, trend-preserving alternative that aligns more closely with observed conditions. While these synthetic datasets have some limitations - particularly in capturing wet extremes of daily maximum humidity - they retain the key climatic signals most relevant to fire behavior modeling, especially for minimum humidity. The resulting products, available across historical and future climate scenarios, provide a robust and accessible resource for wildfire projections, vegetation modeling, and other climate impact analyses, offering improved reliability for long-term planning and risk assessment in California. Their suitability for applications in other sectors has not been investigated.

References

Bryant, B.P. and Westerling, A.L., 2014. Scenarios for future wildfire risk in California: links between changing demography, land use, climate, and wildfire. *Environmetrics*, *25*(6), pp.454-471.

Pierce, D.W., Westerling, A.L. and Oyler, J., 2013. Future humidity trends over the western United States in the CMIP5 global climate models and variable infiltration capacity hydrological modeling system. *Hydrology and Earth System Sciences*, *17*(5), pp.1833-1850.

Preisler, H.K. and Westerling, A.L., 2007. Statistical model for forecasting monthly large wildfire events in western United States. *Journal of Applied Meteorology and Climatology*, *46*(7), pp.1020-1030.

Preisler, H.K., Westerling, A.L., Gebert, K.M., Munoz-Arriola, F. and Holmes, T.P., 2011. Spatially explicit forecasts of large wildland fire probability and suppression costs for California. *International Journal of Wildland Fire*, *20*(4), pp.508-517.

Rahimi, S. 2022. Memo on the Evaluation of Downscaled GCMs Using WRF. https://www.energy.ca.gov/media/10479

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Krantz, W., Bass, B., Thackeray, C., Lin, H., Chen, D., Dennis, E., Collins, E., Lebo, Z. J., Slinskey, E., Graves, S., Biyani, S., Wang, B., Cropper, S., and the UCLA Center for Climate Science Team: An overview of the Western United States Dynamically Downscaled Dataset (WUS-D3), Geosci. Model Dev., 17, 2265–2286, https://doi.org/10.5194/gmd-17-2265-2024, 2024.

Simpson, I.R., McKinnon, K.A., Kennedy, D., Lawrence, D.M., Lehner, F. and Seager, R., 2024. Observed humidity trends in dry regions contradict climate models. *Proceedings of the National Academy of Sciences*, *121*(1), p.e2302480120.

Westerling, A.L. and Bryant, B.P., 2008. Climate change and wildfire in California. *Climatic Change*, *87*(Suppl 1), pp.231-249.

Westerling, A.L., Turner, M.G., Smithwick, E.A., Romme, W.H. and Ryan, M.G., 2011. Continued warming could transform Greater Yellowstone fire regimes by mid-21st century. *Proceedings of the National Academy of Sciences*, *108*(32), pp.13165-13170.

Westerling, A.L., Bryant, B.P., Preisler, H.K., Holmes, T.P., Hidalgo, H.G., Das, T. and Shrestha, S.R., 2011. Climate change and growth scenarios for California wildfire. *Climatic Change*, *109*(Suppl 1), pp.445-463.