Response to Editor

The "paradox" of dryland expansion and Earth greening is an important scientific question. This study wants to study global dryland aridity changes indicated by atmospheric, hydrological, and vegetation observations at meteorological stations. This work has great potential to improve our understanding of arid ecohydrology in climate change. Both two reviewers also gave positive comments on the scientific significance and quality of this work. I generally agree with their judgement. One thing I'd like to raise the authors attention is the underneath root zone. Root zone is a key element in ecohydrology, and root zone dynamic in response to climate change has great impacts on terrestrial ecosystem and hydrological processes. The authors can refer these root zone relevant papers, and add discussions in either Introduction or Discussion parts: (Gao et al., 2014, GRL; Wang-Erlandsson et al., 2016, HESS, Singh et al., 2020, ERL).

Response & actions:

Thank you for the insightful comments and also for your time spent reviewing and processing our manuscript. We have revised this manuscript based on your suggestions and the two reviewers' comments.

Based on your and the reviewers' suggestions, we have elaborated the discussion related to the belowground root zone and referenced relevant papers in the Discussion section:

'The low performance of some flux stations (e.g., shrubland stations), may be related to inadequate modelling of the influence of belowground hydrologic processes. Belowground hydrogeologic properties and groundwater dynamics are difficult to quantify directly through remote sensing or meteorological data. It is thus difficult to capture the effects of subterranean ventilation (López-Ballesteros et al., 2017) and the dynamic relationship between plant root zone and groundwater. Previous studies have shown that the root zone storage capacity (Gao et al., 2014; Wang-Erlandsson et al., 2016; Singh et al., 2020) is important in hydrological processes in drylands and during drought events. Researchers have attempted to estimate root depth and root zone storage capacity (Wang-Erlandsson et al., 2016; Stocker et al., 2023), or to couple drylands' deep-root distribution modules into earth system models (Zhang et al., 2013; Li et al., 2015), and improved the hydrological and ecological prediction (Gao et al., 2014). However, in these approaches, there remain partial limitations such as the dependency on satellite-based ET data (Wang-Erlandsson et al., 2016) containing uncertainty. On the other hand, accurately modelling groundwater dynamics remains limited (Gleeson et al., 2016, 2021). Uncertainties in station-scale groundwater dynamics also affect our understanding of the root-groundwater relationship and groundwater's contribution to ET. Combining drought index at different time scales (e.g., the Standardized Precipitation Evapotranspiration Index (SPEI)) at the regional scale (Secci et al., 2021), and the Gravity Recovery and Climate Experiment (GRACE) based anomalies in terrestrial water storage (Li et al., 2019) can be promising in indirectly representing the groundwater dynamics, but mismatches in spatial scales may still cause errors.' (line 284-302)

Response to Reviewer1

The authors evaluated global changes in dryland aridity using data from meteorological sites during 2003-2019, with the goal of reducing scale-related uncertainty. They obtained a comprehensive understanding of the multifaceted characteristics of these changes and identified a decoupling between atmospheric, hydrological, and vegetation aridity. The manuscript is well-written, and the results are intriguing.
Response:
Thank you for your positive comments. Thank you also for your time spent reviewing our manuscript. We will revise this manuscript based on your insightful suggestions.

Some specific points:
Method: the scale mismatch between the site observations and the gridded data used in the RF model may introduce uncertainty. The authors may want to consider addressing this uncertainty in their modeling practice, or at the very least, discuss it in Section 4.
Response: Thank you for your insightful comments.
Predictor data extractions in our study were at the 500 m scale, which can be matched to the flux observation scale of the stations used in this study. We will discuss possible uncertainties regarding the degree of match between the footprint area of the flux station and the scale of the predictor data used.

Action: elaborated in the discussion section:
'In addition, mismatches between the flux footprints of flux stations and remote sensing data pixels may also cause uncertainty, especially if the flux footprints include considerable spatial heterogeneity (Chu et al., 2021). The 500 m scale of data extraction in this study may have reduced this effect partially, but it may still exist due to the variability of flux footprints across sites. Previous studies have shown that when data are extracted at scales larger than 500 m, the representativeness of the flux footprint area's land cover types can be considerably decreased (Chu et al., 2021). The use of a fixed target area extent for data extraction may bias model-data integration in multi-station level studies. In the future, to reduce the related bias, we should pay more attention to the heterogeneity within the flux footprints of specific flux sites especially in remote sensing data extraction and processing (Walther et al., 2022).' (line274-282)

Figure 2(a): why is the performance of one site exceptionally poor with a negative correlation (Rcorr<0)?
Response: Thank you for your insightful comments.
This site is ES-AMO and the low accuracy of ET simulations at this site may be induced by its hydrogeologic characteristics and groundwater table dynamics (López-Ballesteros et al., 2017). The predictors used in this study may have difficulty explaining the contribution of groundwater dynamics and subterranean ventilation (López-Ballesteros et al., 2017) to ET at the site location. The biological crust at this site (Chamizo et al., 2016) may also control ET by influencing surface soil moisture.

In addition, since our accuracy evaluation is based on the leave-one-site-out cross-validation, the validation accuracy may be relatively low when there are no stations with similar environmental conditions in the training set. The RF model that we finally applied to the weather stations included all stations (i.e., no flux station was left), the accuracy may be improved when applied to weather stations with similar environmental conditions to this flux station.

In the Discussion section, we will elaborate on this issue and clarify possible limitations and uncertainties.

Action: elaborated in the discussion section:
'The low performance of some flux stations (e.g., shrubland stations), may be related to inadequate modelling of the influence of belowground hydrologic processes. Belowground hydrogeologic properties and groundwater dynamics are difficult to quantify directly through remote sensing or meteorological data. It is thus difficult to capture the effects of subterranean ventilation (López-Ballesteros et al., 2017) and the dynamic relationship between plant root zone and groundwater. Previous studies have shown that the root zone storage capacity (Gao et al., 2014; Wang-Erlandsson et al., 2016; Singh et al., 2020) is important in hydrological processes in drylands and during drought events. Researchers have attempted to estimate root depth and root zone storage capacity (Wang-Erlandsson et al., 2016; Stocker et al., 2023), or to couple drylands' deep-root distribution modules into earth system models (Zhang et al., 2013; Li et al., 2015), and improved the hydrological and ecological prediction (Gao et al., 2014). However, in these approaches, there remain partial limitations such as the dependency on satellite-based ET data (Wang-Erlandsson et al., 2016) containing uncertainty. On the other hand, accurately modelling groundwater dynamics remains limited (Gleeson et al., 2016, 2021). Uncertainties in station-scale groundwater dynamics also affect our understanding of the root-groundwater relationship and groundwater's contribution to ET. Combining drought index at different time scales (e.g., the Standardized Precipitation Evapotranspiration Index (SPEI)) at the regional scale (Secci et al., 2021), and the Gravity Recovery and Climate Experiment (GRACE) based anomalies in terrestrial water storage (Li et al., 2019) can be promising in indirectly representing the groundwater dynamics, but mismatches in spatial scales may still cause errors. In addition, our accuracy evaluation was based on the leave-one-station-out cross-validation (Zhang et al., 2021). The validation accuracy may be relatively low when there are no stations with similar environmental conditions in the training set. The RF model that we finally applied to the weather stations included all stations (i.e., no flux station was left), the accuracy can thus be improved a little, especially at weather stations with similar environmental conditions (e.g., shrubland stations) to the previously left flux station in the leave-one-station-out cross-validation.' (line 284-308)

Figure 3: the Antarctic continent could be omitted from the figure to enhance the clarity of the focal information.
Response:
The Antarctic continent will be removed from the figures.

Action: It is removed.

Lines 28-33 and Lines 279-283: the content provided is repetitive. Please rephrase and avoid redundancy.

Response: Thank you for your insightful comments. We will rephrase and avoid redundancy.

Action: Revised in the conclusion section:

'It shows that global drylands' atmospheric, hydrological and ecological aridity changes are inconsistent. Specifically, atmospheric aridity increased and ecological aridity decreased. Changes in hydrologic aridity were not significant in most of the dominant combinations of VPD, LAI, and P-ET. This study highlights the significance to investigate dryland aridity changes using weather station scale data, which can complement previous findings based on coarse-resolution climate reanalysis. It also has the promise of being combined with more station-scale data to provide support for local community's climate change adaptation.'

Response to Reviewer2

This is an interesting study. Combining the station observation and the Random Forest (RF) model, the authors investigated the global dryland aridity changes from 2003 to 2019. They found an evident decoupling between atmospheric, hydrological, and vegetation aridity.

Compared to the prior studies, I think this work creatively assessed global aridity with station data, instead of reanalysis dataset or numeric model outputs. The results are helpful to our understanding of the influence of climate change on the global aridity.

The results are convincing and also highlight the added values of using station scale data to assess dryland change as a complement to the results based on coarse resolution reanalysis data and land surface models.

The manuscript was technically well-written and easy to follow in logic. However, the following minor and technical revisions should be addressed.

Response:
Thank you for your positive comments. Thank you also for your time spent reviewing our manuscript. We will revise this manuscript based on your insightful suggestions.

Lines 38-39

Change to "In the context of global warming, the global dryland is expected to expand due to potential higher atmospheric water demand."

Response: Thank you for your insightful comments.
It will be modified.
Action: It is modified as 'In the context of global warming, the global dryland is expected to expand due to potential higher atmospheric water demand'.

Figure 1 Please add some related references to support your classification of AI levels.

Response:
Thank you for your insightful comments. It will be added. Here, aridity is classified by the index of aridity (AI) as the average annual precipitation divided by potential evapotranspiration (Programme, 1997). Hyperarid: AI < 0.05, Arid: 0.05 < AI < 0.20, Semiarid: 0.20 < AI < 0.50, Dry sub-humid: 0.50 < AI < 0.65, and Humid: AI > 0.65.
Action: the reference is added in Figure 1 caption: AI level classification (Programme, 1997): hyperarid (0 < AI < 0.05), arid (0.05 < AI < 0.2), semiarid (0.2 < AI < 0.5), dry subhumid (0.5 < AI < 0.65).

Lines 126-128 Provide the full name of RMSE. What does the feature importance mean?

Additionally, I recommend adding a detailed introduction to your Random Forest model and leave-one-site-out cross-validation method in the Methodology Section.

Response: Thank you for your insightful comments.
Descriptions of accuracy metrics, feature importance, and random forest models are added to the manuscript to provide more details to the reader. In terms of the leave-one-out cross-validation method, we referred to previous studies (Tramontana et al., 2016; Zhang et al., 2021).

Action: 'the root mean square error (RMSE)' is added.

Added 'The RF model was constructed using the RandomForestRegressor function from the scikit-learn package of Python. The parameter 'n_estimators' was set to 500, and default parameter values were used for the other parameters (Zhao et al., 2019). For the evaluation of model performance, we used a leave-one-station-out cross-validation approach used in previous studies of ET predictions (Tramontana et al., 2016; Zhang et al., 2021; Shi et al., 2022). It is a type of cross-validation approach in which each station's observation is considered as the validation set and the rest stations' observations are considered as the training set. It can help us understand the potential adaptability of the model to new data in the prediction set. Feature importance (IMP) was used to measure the contributions of predictors, and we adopted the permutation importance indices to represent IMP due to their reliability (Díaz-Uriarte and Alvarez de Andrés, 2006; Strobl et al., 2008; Grömping, 2009; Zhang et al., 2021) in RF models.' (line 107-117)

Figure 2 Could you explain the lower Rcorr value of Shrubland in Figure 2c, compared to other land use types?

Response: Thank you for your insightful comments.
The low accuracy of shrubland sites such as ES-AMO may be induced by its hydrogeologic characteristics and groundwater table dynamics (López-Ballesteros et al., 2017), and also associated limitations in the quantifying the belowground hydrological process. The predictors used in this study may have difficulty explaining the contribution of groundwater dynamics and subterranean ventilation (López-Ballesteros et al., 2017) to ET at the site location. The inclusion of cumulative soil water deficit (Giardina et al., 2023) during the growing season and regional-scale drought severity has the potential to better characterize belowground water availability. In addition, in arid shrub ecosystems, the biological crust (Chamizo et al., 2016) may also control ET by influencing surface soil moisture.

In the Discussion section, we will elaborate on this issue and clarify possible limitations and uncertainties.

Action: elaborated in the discussion section:
'The low performance of some flux stations (e.g., shrubland stations), may be related to inadequate modelling of the influence of belowground hydrologic processes. Belowground hydrogeologic properties and groundwater dynamics are difficult to quantify directly through remote sensing or meteorological data. It is thus difficult to capture the effects of subterranean ventilation (López-Ballesteros et al., 2017) and the dynamic relationship between plant root zone and groundwater. Previous studies have shown that the root zone storage capacity (Gao et al., 2014; Wang-Erlandsson et al., 2016; Singh et al., 2020) is important in hydrological processes in drylands and during drought events. Researchers have attempted to estimate root depth and root zone storage capacity (Wang-Erlandsson et al., 2016; Stocker et al., 2023), or to couple drylands' deep-root distribution modules into earth system models (Zhang et al., 2013; Li et al., 2015), and improved the hydrological and ecological prediction (Gao et al., 2014). However, in these approaches, there remain partial limitations such as the dependency on satellite-based ET data (Wang-Erlandsson et al., 2016) containing uncertainty. On the other hand, accurately modelling groundwater dynamics remains limited (Gleeson et al., 2016, 2021). Uncertainties in station-scale groundwater dynamics also affect our understanding of the root-groundwater relationship and groundwater's contribution to ET. Combining drought index at different time scales (e.g., the Standardized Precipitation Evapotranspiration Index (SPEI)) at the regional scale (Secci et al., 2021), and the Gravity Recovery and Climate Experiment (GRACE) based anomalies in terrestrial water storage (Li et al., 2019) can be promising in indirectly representing the groundwater dynamics, but mismatches in spatial scales may still cause errors. In addition, our accuracy evaluation was based on the leave-one-station-out cross-validation (Zhang et al., 2021). The validation accuracy may be relatively low when there are no stations with similar environmental conditions in the training set. The RF model that we finally applied to the weather stations included all stations (i.e., no flux station was left), the accuracy can thus be improved a little, especially at weather stations with similar environmental conditions (e.g., shrubland stations) to the previously left flux station in the leave-one-station-out cross-validation.' (line 284-308)

Lines 136-137 and Figure 3

Why did you separate the entire study period into the two stages (i.e., 2003-2010 and 2011-2019)? Does the 2003-2009 and 2010-2019 work?

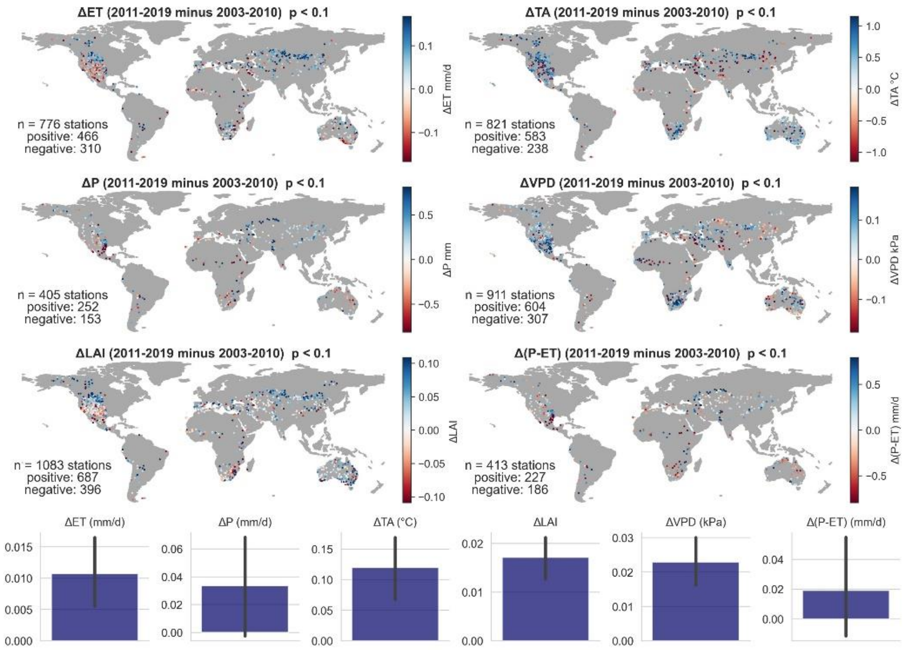
Response:
Thank you for your insightful comments. 2003-2010 and 2011-2019 have similar year lengths, which can somewhat reduce the impact of extreme years on the results and understanding of the period scale. Due to the temporal limitation of MODIS data and predictors derived from it (such as the radiation-related predictor RSDN starts from 2002), the year span of this study is not very long, thus the finding of changes in aridity should be treated with more caution. We will elaborate on the discussion.

Action: elaborated in the discussion section:

‘Compared to previous dryland change studies with decades of span, the period in this study is only 2003-2019 due to the constraint of using MODIS-derived data. We split 2003-2019 into two periods with similar year spans, 2003-2010 and 2011-2019. In this way, it is possible to reduce the effect of extreme years when comparing the differences between the two periods. However, the year spans in this study are not very long compared to studies with longer time series (Lian et al., 2021; Huang et al., 2016), and thus the associated findings should be treated with more caution.’ (line327-332)

I recommend adding a column plot in Figure 3 to summarize the changes in ET, TA, P, VPD, LAI, and P-ET.

Response:
Thank you for your suggestion. We will add column plots to show changes in these variables.
Action: It has been added.



References

Chamizo, S., Cantón, Y., Rodríguez-Caballero, E., and Domingo, F.: Biocrusts positively affect the soil water balance in semiarid ecosystems, Ecohydrology, 9, 1208–1221, https://doi.org/10.1002/eco.1719, 2016.

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N. A., Chen, J., Chen, X., Clark, K., Desai, A. R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J. A., Gough, C. M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J. F., Knox, S. H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J. W., Noormets, A., Novick, K., Oberbauer, S. F., Oechel, W., Oikawa, P., Papuga, S. A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W. L., Richardson, A. D., Russell, E. S., Scott, R. L., Starr, G., Staebler, R., Stoy, P. C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R. C., Suyker, A., Ueyama, M., Vargas, R., Wood, J. D., and Zona, D.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agricultural and Forest Meteorology, 301–302, 108350, https://doi.org/10.1016/j.agrformet.2021.108350, 2021.

Díaz-Uriarte, R. and Alvarez de Andrés, S.: Gene selection and classification of microarray data using random forest, BMC bioinformatics, 7, 1–13, 2006.

Gao, H., Hrachowitz, M., Schymanski, S., Fenicia, F., Sriwongsitanon, N., and Savenije, H.: Climate controls how ecosystems size the root zone storage capacity at catchment scale, Geophysical Research Letters, 41, 7916–7923, 2014.

Giardina, F., Gentine, P., Konings, A. G., Seneviratne, S. I., and Stocker, B. D.: Diagnosing evapotranspiration responses to water deficit across biomes using deep learning, New Phytologist, 240, 968–983, https://doi.org/10.1111/nph.19197, 2023.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geoscience, 9, 161–167, 2016.

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., and Rahman, S.: GMD Perspective: The quest to improve the evaluation of groundwater representation in continental to global scale models, Geoscientific Model Development Discussions, 2021, 1–59, 2021.

Grömping, U.: Variable importance assessment in regression: linear regression versus random forest, The American Statistician, 63, 308–319, 2009.

Huang, J., Yu, H., Guan, X., Wang, G., and Guo, R.: Accelerated dryland expansion under climate change, Nature Clim Change, 6, 166–171, https://doi.org/10.1038/nclimate2837, 2016.

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., and Mukherjee, A.: Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges, Water Resources Research, 55, 7564–7586, 2019.

Li, C., Zhang, C., Luo, G., Chen, X., Maisupova, B., Madaminov, A. A., Han, Q., and Djenbaev, B. M.: Carbon stock and its responses to climate change in C entral A sia, Global change biology, 21, 1951–1967, 2015.

Lian, X., Piao, S., Chen, A., Huntingford, C., Fu, B., Li, L. Z. X., Huang, J., Sheffield, J., Berg, A. M., Keenan, T. F., McVicar, T. R., Wada, Y., Wang, X., Wang, T., Yang, Y., and Roderick, M. L.: Multifaceted characteristics of dryland aridity changes in a warming world, Nat Rev Earth Environ, 2, 232–250, https://doi.org/10.1038/s43017-021-00144-0, 2021.

López-Ballesteros, A., Serrano-Ortiz, P., Kowalski, A. S., Sánchez-Cañete, E. P., Scott, R. L., and Domingo, F.: Subterranean ventilation of allochthonous CO2 governs net CO2 exchange in a semiarid Mediterranean grassland, Agricultural and Forest Meteorology, 234–235, 115–126, https://doi.org/10.1016/j.agrformet.2016.12.021, 2017.

Programme, U. N. E.: World Atlas of Desertification: Second Edition, 1997.

Secci, D., Tanda, M. G., D'Oria, M., Todaro, V., and Fagandini, C.: Impacts of climate change on groundwater droughts by means of standardized indices and regional climate models, Journal of Hydrology, 603, 127154, https://doi.org/10.1016/j.jhydrol.2021.127154, 2021.

Shi, H., Luo, G., Hellwich, O., Xie, M., Zhang, C., Zhang, Y., Wang, Y., Yuan, X., Ma, X., Zhang, W., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Evaluation of water flux predictive models developed using eddy-covariance observations and machine learning: a meta-analysis, Hydrology and Earth System Sciences, 26, 4603–4618, https://doi.org/10.5194/hess-26-4603-2022, 2022.

Singh, C., Wang-Erlandsson, L., Fetzer, I., Rockström, J., and Van Der Ent, R.: Rootzone storage capacity reveals drought coping strategies along rainforest-savanna transitions, Environmental Research Letters, 15, 124021, 2020.

Stocker, B. D., Tumber-Dávila, S. J., Konings, A. G., Anderson, M. C., Hain, C., and Jackson, R. B.: Global patterns of water storage in the rooting zones of vegetation, Nat. Geosci., 16, 250–256, https://doi.org/10.1038/s41561-023-01125-2, 2023.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A.: Conditional variable importance for random forests, BMC Bioinformatics, 9, 307, https://doi.org/10.1186/1471-2105-9-307, 2008.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., Carvalhais, N., Ermida, S. L., Brümmer, C., Schrader, F., Prokushkin, A. S., Panov, A. V., and Jung, M.: Technical note: A view from space on global flux towers by MODIS and Landsat: the FluxnetEO data set, Biogeosciences, 19, 2805–2840, https://doi.org/10.5194/bg-19-2805-2022, 2022.

Wang-Erlandsson, L., Bastiaanssen, W. G., Gao, H., Jägermeyr, J., Senay, G. B., Van Dijk, A. I., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H.: Global root zone storage capacity from satellite-based evaporation, Hydrology and Earth System Sciences, 20, 1459–1481, 2016.

Zhang, C., Li, C., Luo, G., and Chen, X.: Modeling plant structure and its impacts on carbon and water cycles of the Central Asian arid ecosystem in the context of climate change, Ecological Modelling, 267, 158–179, https://doi.org/10.1016/j.ecolmodel.2013.06.008, 2013.

Zhang, C., Luo, G., Hellwich, O., Chen, C., Zhang, W., Xie, M., He, H., Shi, H., and Wang, Y.: A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data from MODIS and flux towers through a machine learning approach, Journal of Hydrology, 603, 127047, https://doi.org/10.1016/j.jhydrol.2021.127047, 2021.

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., and Qiu, G. Y.: Physics-Constrained Machine Learning of Evapotranspiration, Geophysical Research Letters, 46, 14496–14507, https://doi.org/10.1029/2019GL085291, 2019.