**Hydrology and
Earth System
Sciences**

Open Access

Discussions

Global patterns of annual actual evapotranspiration with land-cover type: knowledge gained from a new observation-based database

S. M. Ambrose and S. M. Sterling

Earth Sciences, Dalhousie University, Halifax, Nova Scotia, Canada

12103

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Abstract

The process of evapotranspiration (ET) plays a critical role in the earth system, driving key land-surface processes in the energy, water and carbon cycles. Land-cover (LC) exerts multiple controls on ET, yet the global distribution of ET by LC and the related
5 physical variables are poorly understood. The lack of quantitative understanding of global ET variation with LC begets considerable uncertainties regarding how ET and key land-surface processes will change alongside ongoing anthropogenic LC transformations.

Here we apply statistical analysis and models to a new global ET database to
10 advance our understanding of how annual actual ET varies with LC type. We derive global fields for each LC using linear mixed effect models (LMMs) that use geographical and meteorological variables as possible independent regression variables.

Our inventory of ET observations reveals important gaps in spatial coverage that overlie hotpots of global change. There is a spatial bias of observations towards the
15 mid latitudes, and LCs with large areas in the high latitudes (lakes, wetlands and barren land) are poorly represented. From the distribution of points as well as the uncertainty analysis completed by bootstrapping we identify high priority regions in need of more data collection.

Our analysis of the new database provides new insights into how ET varies globally,
20 providing more robust estimates of global ET rates for a broad range of LC types. Results reveal that different LC types have distinct global patterns of ET. Furthermore, zonal ET means among LCs reveal new patterns: ET rates in low latitudinal bands are more sensitive to LC change than in higher latitude bands; LCs with a higher evaporation component show higher variability of ET at the global scale; and LCs with
25 dispersed rather than contiguous global locations have a higher variability of ET at the global scale.

Results from this study indicate two major advancements are required to improve our ability to predict how ET will vary with global change. First, further collection of ground

12104

truth observations of ET is needed to fill gaps in LC types and spatial location identified in this paper. Second, LC types need to be de-aggregated into finer categories to better characterize ET, to reduce uncertainty and weakened strength to predictor variables, associated by aggregation of heterogeneous LC types into one group; this will require the development of higher-resolution LC databases.

1 Introduction

Evapotranspiration (ET), the land-surface flux of the water cycle, is a critical process in the Earth system as it drives land–atmosphere interactions for three major global cycles, the energy, water, and carbon cycles, directly and indirectly affecting surface temperature, plant productivity, and water availability. Accurate global-scale estimates of ET are thus critical for better understanding of climatological (Shukla and Mintz, 1982), hydrological (Mueller et al., 2013), and carbon interactions (Jasechko et al., 2013). However, estimations of ET made by process-based models are uncertain due to the complexity and nonlinearity of the systems governing ET as well as the lack of reference observations to validate the estimates (Mueller et al., 2011). Thus, independent global spatial fields of ET are needed for validation of Land Surface Models (LSMs), and to increase our understanding of the spatial patterns in the water cycle (Boé and Terray, 2008; Mueller et al., 2011; Seneviratne et al., 2010).

Land-cover (LC) type influences the four major pathways that drive ET, through: (a) water availability, (b) energy availability, (c) photosynthesis rates, and (d) atmospheric moisture gradient (Supplement A; Sterling et al., 2013). LC change alters water availability at the land surface by changing rooting depth, changing soil properties that retain moisture, and by directly removing or adding water to the surface through irrigation, inundation and draining. LC change alters energy availability by changing the albedo and the thermal inertia of the surface. LC change alters photosynthesis rates by changing leaf area, stomatal density, water use efficiency (e.g., C3 to C4 plants)

and nutrient availability. Lastly, LC change alters the atmospheric moisture gradient by altering the surface roughness and therefore surface turbulent exchange.

While ET is challenging both to model and to measure on the ground, there are a range of ET datasets that have advanced our understanding of this important flux: diagnostic observation-based datasets, reanalyses, and uncoupled or coupled LSMs, with each having its own biases and limitations (Mueller et al., 2011, 2013). These datasets are useful in establishing global patterns of ET, and how they vary by climate type, but do not deliberately set out to estimate ET for individual LC types. To date, there is no available database designed to test the response of ET to LC change, particularly one that covers a broad range of LC types. Thus, the characterization of ET rates and patterns among different LC types has remained elusive.

Here we use a new assembly of information on point-based estimates of ET for discrete LC types to gain new insights on how ET varies with a broad range of LC types. We create and employ a database of point-based estimates of annual actual ET rates (ET_OBS) and generate global fields of ET (ET_LMM) from these observations generated with a linear mixed effect model (LMM) using meteorological and geographical predictors. We examine the following questions:

1. What are the patterns in available information from ET observations among LC types?

2. What are the global patterns of annual actual ET for different LC types?

3. What are the patterns of uncertainties for global predictions of ET using a statistical model among LC types?

2 Methods

2.1 GETA 2.0 database

The GETA 2.0 (**G**lobal **ET A**ssembly 2.0) database is a new global-scale dataset of annual actual ET rates classified by LC type (Table 1, Supplement B). The GETA 2.0 database comprises 2363 points across the globe representing 16 LC types (Table 1). GETA 2.0 has improved from its first version used in Sterling et al. (2013) through the addition of over 800 more data points. The ET values in the GETA 2.0 database (ET_OBS) include estimates covering the period 1850 to 2010, with records varying in length from 1 to 107 years. Data are collected with a variety of methods, including eddy covariance, energy balance, soil moisture balance, and water balance methods, over variable scales from plot studies to larger catchments. Criteria for inclusion in the database are that the data points had to be published in government or scientific literature, represent annual actual ET, represent ET for a single LC type, and be representative of a specific location on the planet. This approach follows fundamental work begun by Helmut Lieth in the 1960s (Lieth, 1972), later furthered by Olson (Olson, 1975) and Atjay and coworkers (Ajtay et al., 1979), that determined characteristic net primary productivity (NPP) fluxes for the major ecosystems of the world based on a database of point observations from around the globe.

GETA 2.0 includes both natural and anthropogenic LCs (Table 1). The LC classes were chosen to be closely linked to plant functional types commonly used in LSMs and to follow the IPCC classification scheme (Sterling and Ducharne, 2008). The LC rasters have a five-minute resolution, enabling the identification of individual wetlands and urban areas, as well as major topographic and physical drivers of the local climate.

2.2 Model development

2.2.1 Independent predictors

Predictors that are independent of LC type were used to model global ET fields for each LC type. We used predictors from the NCC 53 year (1948 to 2000), 6 h, meteorological forcing dataset (Ngo-Duc et al., 2005), an elevation dataset (United States Geological Survey, 1997), and latitude and longitude. The annual average values of seven meteorological predictors (Table 2) were extracted from the NCC dataset. The half-century timeframe of NCC captures the same timeframe in which most of the ET observational data were collected. Like other atmospheric forcing datasets, the NCC dataset is designed to describe the overlying meteorology for a variety of LC types at a particular location on the surface, and can be considered independent from LC type. It should be noted, however, that these atmospheric forcing datasets are based upon data from meteorological stations located on the surface that are typically situated on grass plots, so the atmospheric forcing data would be representative more of grass plots than any other LC type. There are 381 GETA 2.0 data points located on large lakes, coastal zones and islands that fall in cells not covered by rasters of the independent predictors. For points that lie on the boundary or within 1° of the NCC raster we manually moved the ET points to the nearest NCC cell from which the information was gathered for the statistical modelling of ET. For the case of large lakes, such as Lake Chad, Lake Superior and the Aral Sea that do not have dataset coverage the elevation and NCC forcing datasets were interpolated using the nearest neighbour value at the coastline for each cell. Small oceanic islands did not have overlying meteorological data, and ET_OBS on such islands and points that were classified generally as forests were not included in LMMing. Thus, of the 2363 GETA data points, 2248 were used to model the global ET_LMM fields.

2.2.2 Modelling global fields of ET

Statistical relationships were identified between ET_OBS and the independent predictors to generate global fields of ET for each LC type. The strength of the relationship between ET_OBS and the independent predictors varies by LC type (Supplement C). Global ET fields were generated at a 1° resolution using generalized LMMing, which allows for the statistical analysis of grouped data (Pinheiro and Bates, 2000). LMM quantifies the variability in ET across and within LC type, while also controlling for independent predictors that may affect this relationship. Instead of fitting a single numerical coefficient in a multiple linear regression model (i.e. a "fixed" effect), the LMM allows coefficients to be normally distributed by group classification (i.e. a "random" effect) where each group-specific effect forms a point on the estimated normal curve. Thus, LMM has fixed and random effects that allow for different rates of predictors for each LC type.

We divided ET_OBS values into groups by LC type and then modelled them in a forward stepwise fashion using the Bayesian Information Criterion (BIC) (Kadane and Lazar, 2004). The best model was created by selecting the predictors that generated the lowest BIC value; the best model identified precipitation and temperature as random effects and shortwave radiation as a fixed effect (Fig. 1). The data were tested against the hypothesis of no spatial autocorrelation ($p < 0.01$) using the Moran's I test statistic. The hypothesis was rejected at $\alpha = 0.05$, so we updated the model to include a spatial correlation based upon the haversine distance between points. The best fit model followed (Eq. 1):

$$\ln(ET) = \mathrm{Intercept_{all}} + \mathrm{Intercept_{LC}} + (\mathrm{TC_{all}} + \mathrm{TC_{LC}}) \cdot T_{air}$$
$$+ (\mathrm{PC_{all}} + \mathrm{PC_{LC}}) \cdot \mathrm{Precip} + \mathrm{SC_{all}} \cdot \mathrm{SW} \tag{1}$$

where $\mathrm{TC_{all}}$ is the fixed intercept for temperature, $\mathrm{TC_{LC}}$ is the random effect for temperature, T_{air} is air temperature, $\mathrm{PC_{all}}$ is the fixed effect for precipitation, $\mathrm{PC_{LC}}$ is the random effect for precipitation, Precip is precipitation, and $\mathrm{SC_{all}}$ is the fixed effect

for shortwave radiation, and SW is short wave radiations; all of these parameters are normalized to z-score values.

An observed strong correlation between the random effects of precipitation and temperature (Fig. 1) was investigated to ensure there was not a problem with the model fit by fitting linear models for each LC individually with only temperature, precipitation and shortwave radiation as predictors; no violations of assumptions were found in the examination of the residuals and the resultant relationships.

Next, we masked the global fields of ET that were generated by the models to the 5 min cells in which the LCs appear. This step changed the 1° ET predictions to 5 min resolution, resulting in a finer resolution located only at their specified LCs (ET_LMM). An implicit assumption here is that the major drivers of ET at the global scale that are not encompassed in the determination of the LC location are changing at a spatial rate that can be represented by a 1° grid (e.g. insolation, precipitation, relative humidity), and that the finer-scale variations in topography and soils are captured in the 5 min LC rasters themselves.

The 5 min presence/absence LC rasters used to mask ET_LMM were derived from Ramankutty and Foley (1999) and Sterling and Ducharne (2008), except for the tree plantation raster which was derived from Erb et al. (2007). We converted percent cover to presence/absence at the 5 min resolution of tree plantations by preserving the area (Kröger, 2012), assuming a linear tree plantation expansion rate between 1990 and 2010, with an estimate of 221.1955 million hectares of tree plantations globally. For all LCs, the rasters used to mask ET_LMM fields did not overlie neatly with the ET_OBS points, as there are cases where ET_OBS measurements were made in cells in which the LC measured was not the dominant LC. 58 % of ET_OBS did not align directly with the overlying LC rasters. Because these points represent a finer resolution of LC not captured in the 5 min rasters, rather than an error in LC classification they remained in the model.

To determine confidence intervals for both the random effect slopes by LC and to generate a map of uncertainty we bootstrapped the raw data. Bootstrapping randomly

samples the raw ET values within groups to simulate possible combinations and then fits a model (Efron and Tibshirani, 1994). This was completed 1000 times to generate a distribution of possible values. We generated the distribution of estimates and the predicted ET rates 95 % confidence interval by extracting the 0.025th and 0.975th quantiles of the bootstrapped data predictions.

3   Results and discussion

3.1   Patterns in available information on annual ET

There is a marked variation of coverage of ET observations (ET_OBS) both spatially and by LC type (Figs. 2 and 3); some LCs and regions in the globe have much fewer ET observations than others. Most ET observations are in forested environments (57.5 % of observations, Fig. 2), and of these most are in evergreen broadleaf and evergreen needle leaf forests; the exception is deciduous needle leaf forests, which has the fewest ET measurements of all LC types and has a limited global extent, predominantly in Siberia. Barren lands and savannah have the fewest ET estimates next to deciduous needle leaf forest, followed by urban lands. As a group, anthropogenic LCs have lower ET coverage than the natural LCs (Fig. 2).

The spatial distribution of ET observations varies markedly across the globe (Fig. 3). Few data points are found in Central Asia and Western Africa. Western Europe and the United States have the densest coverage of ET measurements, although these areas of highest density do not necessarily imply sufficient spatial coverage of ET measurements.

Regions with lower ET_OBS coverage intersect key global hotspots. South East Asia is a hotspot for change in ET with LC change (Boisier et al., 2014; Sterling et al., 2013), for correlation of summer temperature and ET (Seneviratne et al., 2006) and for high threat to Human Water Security (HWS) (Vorosmarty et al., 2010). Equatorial Africa is a hotspot of ET change due to LC change (Boisier et al., 2014; Sterling et al.,

2013), a high risk to HWS (Vorosmarty et al., 2010), and projected early temperature departure from atmospheric CO_2 increase (Mora et al., 2013). Other areas with poor ET coverage that intersect areas of high threat to HWS and change in ET from LC change include western South America, India, Eastern China and Afghanistan, Western Australia, Central America, East Africa, European Russia, Western Asia and Southern Europe.

Our statistical model predicts that for the LCs where temperature greatly contributes to higher ET_LMM values, precipitation does not and vice versa (Fig. 1, Table 3). This relationship follows the theoretical moisture vs. energy-limited status of LC types (e.g. Creed et al., 2014). For example, the ET_LMM rates for barren lands (BAR, often moisture limited) increase with increasing precipitation, with almost no change with increasing temperature. Conversely, ET_LMM rates for wetlands and irrigated agriculture (by definition energy limited) increase with temperature, with almost no change with increasing precipitation. ET reacts to a more equal combination of precipitation and temperature drivers for forests, savannah, non-irrigated agriculture, tree plantations, grazing lands and urban lands.

3.2   Global ET means for individual LCs

Our analysis reveals that the order in which terrestrial biomes have the highest to lowest ET rates follows the established order of biome NPP ranking (Lieth, 1975; Olson et al., 1983; Saugier et al., 2001). The mean annual ET rates (for both ET_OBS and ET_LMM) are highest for evergreen broadleaf forest, irrigated croplands, and wetlands and are lowest for barren land (Fig. 4, Table 1). The majority of ET_OBS lie between 0.3 and 1.5 m yr^{-1} (Fig. 4), with the observed mean equal to 0.84 m yr^{-1} and the median equal to 0.69 m yr^{-1}.

The data show that many LC types have significantly different global annual average observed ET (ET_OBS) from other LCs (Table 2). Of the 120 different comparisons of ET_OBS LC types, 70 (58.3 %) are significantly different at $\alpha = 0.05$ using Tukey's mean comparison test (Table 4). Non-irrigated agriculture and deciduous broadleaf

forests are the LCs with the fewest significant differences from other LC types, these LCs have ET values in the middle range of ET (0.25 to 1.0 m yr^{-1}), with means between 0.6 to 0.7 m yr^{-1}. Lakes have a significantly higher ET rate than for all other LC types.

However, the ET comparisons based on ET_OBS data are limited by the spatial coverage of the observations. Most ET observations are located in mid-latitudes, and LCs with large areas in higher latitudes (e.g. lakes, wetlands and barren lands) are more poorly-represented by the current ET observations. As a result, the mean global average ET for these LCs derived from ET_OBS is too high. For these LCs, ET_OBS is higher than the ET_LMM mean (Table 1, Fig. 4), as expected because ET_LMM covers the high-latitudes. The difference in maximum value for the ET_OBS and ET_LMM is greatest for lakes, barren lands and wetlands.

For other LCs, the means of ET_OBS and ET_LMM are similar, such as for evergreen broadleaf forest, deciduous broadleaf forests, savannah, shrub land, irrigated agriculture and urban lands; the location of these LCs tend to be in the mid- to low- latitudes for which there is better representation in ET_OBS. In general, the model predictions (ET_LMM) are well within one standard deviation of ET_OBS for most LCs (Table 1, Fig. 4).

The GETA 2.0 ET estimates for individual LCs correspond well with another estimate (Rockström et al., 1999) that summarized mean ET for LC types, although comparisons are challenging because of lack in congruency in LC classes among studies. For the four LC types that are congruent with GETA 2.0 types the Rockström et al. (1999) mean annual ET estimates fall within the ranges of ET_OBS and ET_LMM. Of interest is the difference in Amazon ET patterns from our results and those from a summary of flux tower data by Jung et al. (2010). Jung et al. (2010) project that ET is lower for wetlands than topical forest in the Amazon, while our findings show wetland ET as higher than the surrounding tropical forest in this region (Fig. 5). The GETA 2.0 results appear to be more accurate because wetland ET rates should be higher than the surrounding forest based on a comparison with NPP; NPP estimates are greater for wetlands than tropical

12113

forests (Amthor and Group, 1998; Lieth, 1975) and there is additional direct evaporation expected from wetlands in addition to the transpiration captured in the wetland NPP.

Global ET_LMM fields project a cumulative total ET (TET) of 70 600 km^3 yr^{-1} for a globe covered with natural vegetation (Fig. 5a and b) and the overlying wetlands. This value coheres with the range of published estimates of TET (73 000 km^3 yr^{-1}, Arora, 2001; 71 000 km^3 yr^{-1}, Baumgartner et al., 1975; 72 900 km^3 yr^{-1}, Berner and Berner, 1987; 60 000 to 85 000 km^3 yr^{-1}, Haddeland et al., 2011; 62 800 km^3 yr^{-1}, Mu et al., 2011; 75 000 km^3 yr^{-1}, Oki, 1999; 56 000 to 84 000 km^3 yr^{-1}, Rockström et al., 1999; and 64 500 to 72 000 km^3 yr^{-1}, Mueller et al., 2013). Our analysis includes major lakes, not included in many other analyses, which may explain why our estimate lies on the higher end of the range of projected TET.

3.3 Zonal patterns of ET

The modelled ET (ET_LMM) results show that different LC types have different zonal patterns of ET (Fig. 6). LCs differ in rates as well as the shape and variability of their zonal ET means. Results indicate that lakes have the highest mean ET rates at almost all latitude points, while deciduous needle leaf forest has the lowest (Table 1, Fig. 6). Of LCs that extend across majority of the latitudes, barren lands and shrublands have the lowest values.

The zonal plots show that the expected response of ET to a particular LC change is not the same across all latitudinal bands. For example, results show that conversion of wetlands to grazing land will decrease annual ET at all latitudes, except for above 60° N; however, the difference is most significant in the mid-latitudes around 25° N and 25° S (Fig. 6). Results suggest that a change from evergreen broadleaf forest to urban lands would cause the greatest reductions in ET in the latitudinal band between about 20° S to 20° N as compared with higher latitude bands. A change from shrub land to lakes would cause the greatest increases in ET between about 20° S to 20° N. Changing barren land to irrigated agriculture at about 45° S likely will cause a smaller increase in

12114

ET then a change at 25° N latitude (Fig. 6). Thus, the zonal plots summarize the likely ET changes resulting from specific LC changes in a particular latitude band.

The most common zonal pattern of ET shows a higher mean ET rate around the equator that declines with increasing latitude but the shape of this relation varies with LC type. The zonal patterns reveal that LCs with highest available surface water (evergreen broadleaf forest, lakes and wetlands) also have the largest increases in ET with latitude towards the equator (Fig. 6). In contrast, LCs with the lowest amount of available surface water (grasslands, shrub lands and barren land) have the smallest increases in ET with latitude towards the equator. The common zonal ET shape is consistent with another study that showed for ET rates for all LCs lumped together peaking at about 0° latitude (Zeng et al., 2012).

Results suggest that the global-scale pattern of a LC location, whether dispersed or contiguous, impacts zonal ET variability. Results show that LCs that have more contiguous global locations (e.g. non-irrigated croplands, deciduous broadleaf forests and shrub lands, Fig. 5) have smoother zonal curves with less variability among latitude bands. In contrast, LCs with dispersed global locations (e.g. lakes, irrigated agriculture, tree plantations, and urban lands, Fig. 5) have more jagged curves (Fig. 6), reflecting higher variability in mean zonal ET in adjacent latitude bands. Thus LCs with larger contiguous areal locations are expected to have more gradual changes in climate among adjacent cells, and more direct feedback over their overlying meteorology than do LCs with dispersed locations. Zonal ET_LMM confidence intervals are largest around the equator and 20–30° N, corresponding with the latitudinal belt that includes a large range of climate types (e.g. deserts in Africa and tropical forests in SE Asia).

Results also indicate differences in global scale ET variability can also be explained by relative roles of vegetation among LC types. LCs for which ET is dominated by one of energy or moisture limitations (barren land, wetlands, and lakes, Fig. 1) have more variability in their zonal patterns with wider confidence intervals around the mean (Fig. 6). In contrast, LCs for which ET is governed through a more equal combination of water availability and energy (such as evergreen needle leaf forest, savannah, and

12115

non-irrigated cropland, Fig. 1) can be considered to be more vegetation-dominated; these LCs have lower variability in their zonal patterns with smaller confidence intervals around the mean (Fig. 6). The first group of LCs are expected to have a greater role of evaporation in ET, as these LCs have more open water and bare soil; thus, a possible explanation for higher ET variability of these LCs is that they can occur in a much wider variety of climates than the vegetation-dominated LCs, and are therefore exposed to a larger range of ET drivers. Indeed, we observe that the ranges of the independent predictors (shortwave radiation, precipitation and air temperature), normalized by the LC area, are greatest for lakes, followed by urban areas and wetlands, and that barren lands have the largest temperature and shortwave radiation range of all the LCs. For the "vegetation-dominated" LCs, the lower variability in global ET can be explained by a smaller range of climate conditions experienced per unit area of the LC, and thus these vegetation dominated LCs are exposed to a lower range of ET drivers. These relationships imply that LCs dominated by energy or moisture limitations will have more uncertainty in their projections with a statistical model than from vegetation-dominated LCs with more balanced energy and moisture limitations to ET. An exception to this observation is that evergreen broadleaf forest (Fig. 6) which is governed by a balance of both energy and moisture limitations also has a wide confidence interval; the higher than expected variability of evergreen broadleaf forest may be due to its very high ET rates or to its lower percentage of interception compared to other forest types (following Miralles et al., 2010). Further research using more refined LC classes is needed to examine these hypotheses.

3.4 Uncertainties in ET for individual LC types

Confidence intervals generated using the bootstrapped models indicate that the largest uncertainties in the ET_LMM predictions lie in Africa and northern latitudes (Fig. 7). These uncertainties qualitatively agree with the density of points generated from ET_OBS in that the lowest uncertainties are located in areas with the most points, such as North America and Europe (Fig. 7), and the largest uncertainties lie in predictions of

12116

Africa, Southern Asia, South America and far northern locations. In terms of LC type, the largest uncertainties are associated with barren lands and wetlands (Fig. 7).

Reliability of ET_LMM predictions also varies with LC type due to varying levels of extrapolation in the model resulting from predictors being used outside of the ET_OBS range. We have mapped the regions of coverage for each LC type to identify the locations the ET predictions are less certain (Supplement D); some of the LCs have poorer coverage of the independent predictors by ET_OBS, leading to less accurate predictions in ET_LMM. In particular, barren land is not well covered by the predictor ranges, resulting in less certain ET predictions outside of the predictor ranges in Northern Africa, Tibet, the Arctic and the Altiplano. These maps also serve to identify where further data collection is needed so that ET fluxes of the LCs can be better understood.

We find that the uncertainty in global modelled annual average values of ET (ET_LMM) varies with LC type and global location. LCs with the highest uncertainty in ET_LMM projection have dispersed locations (Fig. 5), lower point and density of coverage (Figs. 2 and 3), larger areas with climate conditions outside the ET_OBS range (Supplement D), and have a greater role of evaporation in ET. In contrast, LCs with the lowest uncertainty have more contiguous locations, higher point and density of coverage, smaller areas with climate conditions outside the ET_OBS range, and are vegetation-dominated. For example, lakes and barren lands have a relatively high uncertainty in ET predictions, associated with decoupling of LC location from climate variables (i.e. more possible climates in which the LC can occur), aggregation of disparate sub-units which would cause global-scale heterogeneity in the data, and extrapolation of ET_LMM beyond range of predictor variables, associated with gaps in measurements in key locations (particularly barren land) (Supplement D).

ET in ocean archipelagos is not well represented in the analysis, due to gaps in coverage of the independent predictors. Advances in atmospheric forcing datasets to cover ocean archipelagos are needed to address this limitation. Future work is also

recommended to extend this analysis to cover sub-annual variation, finer spatial scales, and the impact of successional stages on ET.

4 Conclusions

Our assembly and statistical analysis of the novel GETA 2.0 database provide a powerful new characterization of global ET patterns for a broad range of LCs, building upon existing knowledge in other databases that do not classify ET by LC type. Our results show that LCs have distinctly different means and zonal patterns. Lakes have the highest ET rates across all latitude bands, although the high latitudes are particularly uncertain.

The global fields of ET produced here (ET_LMM) are useful as a reference for process-based model estimates of ET because they are derived from statistical modelling of observations, and do not rely on the same assumptions used in deterministic land surface and climate models. Furthermore, information presented here on ranges of ET observations for individual LC types (Table 1, Fig. 3) can be used to support decisions on whether modeled ET outliers can be excluded based on physical considerations (e.g. McCabe et al., 2008).

An interesting finding is that ET is more sensitive to LC change in some latitude bands than others. Changes to ET from LC change are expected to be largest in the tropical latitude bands, particularly resulting from changes between LCs with high water availability and low water variability. This finding implies that global land use planning must be particularly careful in the tropics because of the possibly heightened possible impacts to the earth system in these latitude bands.

Our analysis suggests two fundamental LC characteristics affect global ET variability. Contiguous LCs display smaller ET variability than dispersed LCs. And LCs that have a heightened role of evaporation in ET have higher variability in global ET than vegetation-dominated LCs. More work to further explore these hypotheses is needed to advance our ability to predict changes to ET with global change.

Our inventory of annual actual ET observations reveals regions and LC types where estimates are most lacking, and many of these are within key global change hotspots in the Arctic, Africa and central Asia. Some LCs have much sparser ET observations than others. Anthropogenic LCs, including grazing land, non-irrigated croplands, and urban areas, emerge are among the most poorly represented LCs by ET observations and we recommend that anthropogenic LCs be targeted for future monitoring of ET. Many anthropogenic LCs also face higher heterogeneity of overlying meteorological conditions within their LC type, related to a greater role of evaporation in ET (e.g. reservoirs), increased spatial dispersion (e.g. urban areas), or poorer density of coverage (e.g. grazing), or a combination of these factors (e.g. irrigated cropland). Thus it is particularly challenging to reliably project ET in anthropogenic LCs using statistical models at the global scale. An added challenge is that anthropogenic LCs are also difficult to parameterize in process-based models because of their heterogeneity. As LC change continues, anthropogenic LCs will play an increasingly significant role in overall global ET rates and advances in modelling and observation networks of ET in human dominated areas are needed. Division of LCs into subclasses for future modelling will help to reduce the heterogeneity needed to better define characteristic the land surface fluxes in these LCs.

Resolution thus is an important characteristic of any global projection of ET. Local high and low values of ET cancel out with coarser spatial resolutions leading to underestimation of ET gradients within continents (Mueller et al., 2011). The relatively fine resolution is an advantage of GETA 2.0 as it captures local features important in terms of climate and water budgets such as individual lakes and wetlands. However, to the extent that the ET_LMM generated here represents only a single climate within a grid cell, the range/variability in LC types and ET rates will always be lower in the model than in the real world, where diverse microclimates can harbour a greater variety of ET rates and LC types. A finer resolution scheme that accounts for more heterogeneity will allow for improved representation of the diversity of microclimates but will inevitably fall short of representing all possible microclimates and LCs.

Issues related to LC change will grow in complexity, as our land surface accumulates a more complicated history of varying types of LC change all while adjusting to changes in overlying meteorology. Continued advances in understanding how ET varies spatially and with LC type are necessary to improve predictions and mitigation actions for the future. The findings here on characteristic ET rates and global patterns with LC type improve our quantitative understanding of the spatial patterns in the water cycle and how ET will change with ongoing anthropogenic transformations; this fundamental information is needed for us to better understand how the earth's energy balance, carbon cycle, and water cycle will respond to global change.

5   List of Supplement

A. Conceptual model of common pathways through which LC change affects ET.

B. The GETA 2.0 Database, all points used in the modelling of ET, as well as an addition 13 points not used in modelling of ET_LMM because of a general forested classification (FFF). References for each estimate are included.

C. Scatter plots of annual actual ET (ET_OBS) and independent predictors (Table 2) for each LC type.

D. Global plots of areas where ET_LMM projections are outside the ranges of independent predictors captured by ET_OBS.

The Supplement related to this article is available online at doi:10.5194/hessd-11-12103-2014-supplement.

References

Ajtay, G. L., Ketner, P., and Duvigneaud, P.: Terrestrial primary production and phytomass, in: The Global Carbon Cycle, edited by: Bolin, B., Degens, E., Kempe, S., and Ketner, P., Wiley, New York, 129–181, 1979.

Amthor, J. and Group, E. W.: Terrestrial Ecosystem Responses to Global Change: a Research Strategy, Technical Report, 37 pp., Oak Ridge Natl. Lab., Oak Ridge, Tenn, 1998.

Arora, V. K.: Streamflow simulations for continental-scale river basins in a global atmospheric general circulation model, Adv. Water Resour., 24, 775–791, 2001.

Baumgartner, A., Reichel, E., and Lee, R.: World Water Balance?: Mean Annual Global, Continental and Maritime Precipitation, Evaporation and Run-off, Elsevier Scientific Publishing Co., Amsterdam, 1975.

Boé, J. and Terray, L.: Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change, Geophys. Res. Lett., 35, L05702, doi:10.1029/2007GL032417, 2008.

Boisier, J. P., de Noblet-Ducoudré, N., and Ciais, P.: Historical land-use induced evapotranspiration changes estimated from present-day observations and reconstructed land-cover maps, Hydrol. Earth Syst. Sci. Discuss., 11, 2045–2089, doi:10.5194/hessd-11-2045-2014, 2014.

Creed, I. F., Spargo, A. T., Jones, J. A., Buttle, J. M., Adams, M. B., Beall, F. D., Booth, E., Campbell, J., Clow, D., Elder, K., Green, M. B., Grimm, N. B., Miniat, C., Ramlal, P., Saha, A., Sebestyen, S., Spittlehouse, D., Sterling, S., Williams, M. W., Winkler, R., and Yao, H.: Changing forest water yields in response to climate warming: results from long-term experimental watershed sites across North America, Global Change Biol., 20, 3191–3208, doi:10.1111/gcb.12615, 2014.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap, CRC Press, 1994.

Erb, K.-H., Gaube, V., Krausmann, F., Plutzar, C., Bondeau, A., and Haberl, H.: A comprehensive global 5 min resolution land-use data set for the year 2000 consistent with national census data, J. Land Use Sci., 2, 191–224, doi:10.1080/17474230701622981, 2007.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance: setup and first results, J. Hydrometeorol., 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J.: Terrestrial water fluxes dominated by transpiration, Nature, 496, 347–50, doi:10.1038/nature11983, 2013.

Kröger, M.: Global tree plantation expansion: a review. ICAS Review Paper Series No. 3, Published jointly by Initiatives in Critical Agrarian Studies (ICAS), Land Deal Politics Initiative (LDPI) and Transnational Institute (TNI), 2012.

Lieth, H.: Modelling the primary productivity of the world, Nat. Resour., 8, 5–10, 1972.

Lieth, H.: Primary production of the major vegetation units of the world, in: Primary Productivity of the Biosphere, Springer, Berlin Heidelberg, 203–215, 1975.

McCabe, M. F., Wood, E. F., Wójcik, R., Pan, M., Sheffield, J., Gao, H., and Su, H.: Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies, Remote Sens. Environ., 112, 430–444, doi:10.1016/j.rse.2007.03.027, 2008.

Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., and Dolman, A. J.: Global canopy interception from satellite observations, J. Geophys. Res., 115, D16122, doi:10.1029/2009JD013530, 2010.

Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser, L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giuseffi, L. M., and Giambelluca, T. W.: The projected timing of climate departure from recent variability, Nature, 502, 183–187, doi:10.1038/nature12540, 2013.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R.,

Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013.

Ngo-Duc, T., Polcher, J., and Laval, K.: A 53-year forcing data set for land surface models, J. Geophys. Res. Atmos., 110, D06116, doi:10.1029/2004JD005434, 2005.

Oki, T.: The global water cycle, in: Global Energy and Water Cycles, edited by: Browning, K. and Gurney, R., Cambridge University Press, 10–27, 1999.

Olson, J. S.: Productivity of forest ecosystems, in: Productivity of World Ecosystems, National Academy of Sciences, Seattle, 33–43, 1975.

Olson, J. S., Watts, J. A., and Allison, L. J.: Carbon in live vegetation of major world ecosystems, Office of Energy Research, US Department of Energy, Washington DC, 1–152, 1983.

Pinheiro, J. C. and Bates, D. M.: Mixed-Effects Models in S and S-PLUS, Springer, New York, 2000.

Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: croplands from 1700 to 1992, Global Biogeochem. Cy., 13, 997–1027, doi:10.1029/1999GB900046, 1999.

Rockström, J., Gordon, L., Folke, C., Falkenmark, M., and Engwall, M.: Linkages among water vapor flows, food production, and terrestrial ecosystem services, Conserv. Ecol., 3, 5, 1999.

Saugier, B., Roy, J., and Mooney, H. A.: Estimations of global terrestrial productivity: converging towards a single number?, in: Terrestrial Global Productivity, Academic Press, San Diego, 543–558, 2001.

Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land–atmosphere coupling and climate change in Europe, Nature, 443, 205–209, doi:10.1038/nature05095, 2006.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: a review, Earth-Sci. Rev., 99, 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Shukla, J. and Mintz, Y.: Influence of land-surface evapotranspiration on the Earth's climate, Science, 215, 1498–1501, doi:10.1126/science.215.4539.1498, 1982.

Sterling, S. M. and Ducharne, A.: Comprehensive data set of global land cover change for land surface model applications, Global Biogeochem. Cy., 22, GB3017, doi:10.1029/2007GB002959, 2008.

Sterling, S. M., Ducharne, A., and Polcher, J.: The impact of global land-cover change on the terrestrial water cycle, Nat. Clim. Change, 3, 385–390, 2013.

United States Geological Survey: Earth Resources Observation and Science Center, available at: http://eros.usgs.gov/elevation-products last access: 4 September 2014, 1997.

Vorosmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, Nature, 467, 555–561, doi:10.1038/nature09440, 2010.

Zeng, Z., Piao, S., Lin, X., Yin, G., Peng, S., Ciais, P., and Myneni, R. B.: Global evapotranspiration over the past three decades: estimation based on the water balance equation combined with empirical models, Environ. Res. Lett., 7, 014026, doi:10.1088/1748-9326/7/1/014026, 2012.

Table 1. LC type and annual actual ET estimation ($m\,yr^{-1}$). Type refers to LC type, whether natural (N) or anthropogenic (A). ET_LMM refers to ET statistics derived from the ET_LMM method. ET_OBS refers to ET statistics derived from the ET_OBS.

LC type	Symbol	Type	ET_LMM ($m\,yr^{-1}$)				ET_OBS ($m\,yr^{-1}$)				Difference of Mean ET (ET_OBS-ET_LMM)
			Max	Range	Mean	Std.	Max	Range	Mean	Std.	
Evergreen broadleaf forest	EBF	N	1.56	1.22	1.21	0.16	3.28	3.15	1.2	0.39	0.01
Deciduous broadleaf forest	DBF	N	1.19	0.84	0.75	0.2	2.41	2.26	0.71	0.35	0.04
Evergreen needle leaf forest	ENF	N	1.34	1.2	0.39	0.16	1.5	1.41	0.56	0.27	−0.17
Deciduous needle leaf forest	DNF	N	0.6	0.44	0.28	0.05	1.95	1.84	0.47	0.47	−0.19
Mixed forest[1]	MXF	N	1.39	1.26	0.34	0.18	2.84	2.66	0.66	0.35	−0.32
Savannah	SAV	N	1.31	1.15	0.78	0.24	3.18	2.91	0.88	0.48	−0.1
Grassland	GRS	N	1.27	1.19	0.42	0.15	2.24	2.23	0.58	0.42	−0.16
Shrubland[2]	SHR	N	0.93	0.92	0.31	0.13	0.97	0.9	0.39	0.22	−0.08
Barren land	BAR	N	1.04	1.04	0.07	0.07	1.67	1.66	0.32	0.3	−0.25
Wetlands	WTL	N	1.94	1.82	0.83	0.51	4.01	3.81	1.06	0.64	−0.23
Lakes and Resevoirs[3]	LAK	A/N	2.59	2.46	0.56	0.39	3.58	3.45	1.61	0.74	−1.05
Irrigated cropland[4]	CRI	A	1.5	1.11	0.93	0.25	4.6	4.3	1.14	0.79	−0.21
Non-irrigated cropland[5]	CRN	A	1.45	1.45	0.65	0.3	1.83	1.66	0.62	0.35	0.03
Tree plantations	TPL	A	1.37	1.17	0.67	0.37	2.5	2.37	0.83	0.4	−0.16
Grazing	GRZ	A	1.66	1.66	0.66	0.32	2.69	2.57	0.77	0.52	−0.11
Urban and built-up	HMO	A	1.18	1.13	0.48	0.17	1.13	1.02	0.52	0.23	−0.04

[1] These are forest areas with mixed species (specified).
[2] Combined closed shrubland and open shrubland.
[3] Includes both lakes and inundated lands, grouped together because of their shared properties.
[4] Includes cells that are dominated by irrigated cropland, but not all cells containing cropland.
[5] The cropland dataset contains locations in desert areas which are likely erroneous, unless irrigated. We did not remove these cells in our global ET field generation. CRN was modified to not include irrigated cropland cells.

Table 2. Independent variables tested as model predictors of ET.

Variable Name	Description	Units	Source
Precipitation	Combined rainfall and snowfall rate	$kg\,m^{-1}$	Ngo-Duc et al. (2005)
T_{air}	Near surface air temperature at 2 m	K	Ngo-Duc et al. (2005)
Q_{air}	Near surface specific humidity at 2 m	$kg\,kg^{-1}$	Ngo-Duc et al. (2005)
Wind	Near surface wind speed at 10 m	$M\,s^{-1}$	Ngo-Duc et al. (2005)
P_{surf}	Surface Pressure	Pa	Ngo-Duc et al. (2005)
Sw_{down}	Surface incident shortwave radiation	$W\,m^2$	Ngo-Duc et al. (2005)
Lw_{down}	Surface incident long wave radiation	$W\,m^2$	Ngo-Duc et al. (2005)
Elevation	Land elevation	m	USGS (2013)
Lat	Latitude	$1° \times 1°$	NA
Long	Longitude	$1° \times 1°$	NA

Table 3. Regression coefficients estimated using a linear mixed model with spatial correlation. These estimates are for the log ET amounts across the globe. Rasters used for prediction have been standardized and transformed to the log scale so that the slopes are directly comparable between predictors and within LCs.

	Intercept	Random Effect T_{air}	Precip	Fixed Effect SW
All	0.64	1.19	1.2	1.14
EBF	1.2	0.95	1.00	NA
DBF	1.04	0.97	0.93	NA
ENF	1.07	1.03	1.01	NA
DNF	1.06	1.04	0.96	NA
MXF	0.99	1.00	1.04	NA
SAV	0.94	1.00	1.00	NA
GRS	0.84	0.92	1.09	NA
SHR	0.71	0.89	1.11	NA
BAR	0.55	0.85	1.44	NA
WTL	1.3	1.15	0.82	NA
LAK	1.54	1.17	0.8	NA
CRI	1.26	1.06	0.85	NA
CRN	0.98	0.99	1.03	NA
TPL	1.07	1.03	0.95	NA
GRZ	1.03	0.97	1.08	NA
HMO	0.83	1.01	1.02	NA

Table 4. Tukey Mean comparison test of differing LC ET_OBS means. Significance at alpha = 0.05 level is noted by bold typeface.

|	EBF	DBF	ENF	DNF	MXF	SAV	GRS	SHR	BAR	WTL	LAK	CRI	CRN	TPL	GRZ	HMO
EBF																
DBF	**0.00**															
ENF	**0.00**	**0.00**														
DNF	**0.00**	0.57	1.00													
MXF	**0.00**	1.00	0.55	0.90												
SAV	**0.00**	0.62	**0.00**	0.05	0.21											
GRS	**0.00**	0.33	1.00	1.00	0.99	**0.01**										
SHR	**0.00**	**0.00**	0.11	1.00	**0.00**	**0.00**	0.14									
BAR	**0.00**	**0.00**	0.05	1.00	**0.00**	**0.00**	0.05	1.00								
WTL	0.36	**0.00**	**0.00**	**0.00**	**0.00**	0.62	**0.00**	**0.00**	**0.00**							
LAK	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**						
CRI	1.00	**0.00**	**0.00**	**0.00**	**0.00**	0.10	**0.00**	**0.00**	**0.00**	1.00	**0.00**					
CRN	**0.00**	0.98	1.00	0.99	1.00	0.18	1.00	0.15	**0.05**	**0.00**	**0.00**	**0.00**				
TPL	**0.00**	0.34	**0.00**	0.05	0.04	1.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.10			
GRZ	**0.00**	1.00	**0.01**	0.31	0.89	1.00	0.17	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.80	1.00		
HMO	**0.00**	0.42	1.00	1.00	0.92	**0.02**	1.00	0.98	0.74	**0.00**	**0.00**	**0.00**	1.00	**0.01**	0.19	
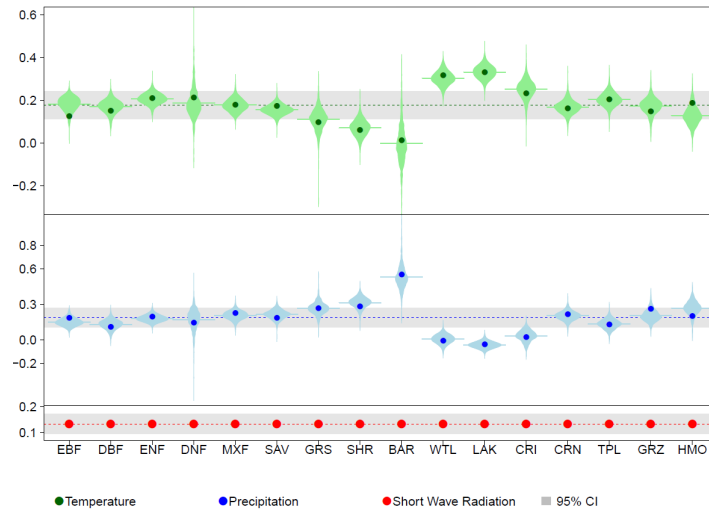
Figure 1. Linear mixed modelling slope by predictor. Slopes of two random effects – air temperature (blue) and precipitation (green) – and fixed slope – shortwave radiation (red). The fixed effect is displayed by the dotted line, with the additional random effect shown by the dot. Lighter shading indicates the bootstrapped distribution, where the horizontal line is the mean of the bootstrapped distribution.
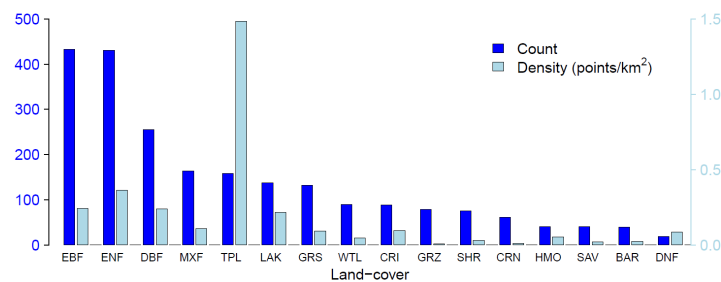
Figure 2. Count and density of ET measurements (ET_OBS) by LC. Count of points in the GETA database for each LC shown in dark blue using the left axis, and the density of points for each land-cover (light blue) calculated per km^2 on the right axis ($\times 10^{-4}$).
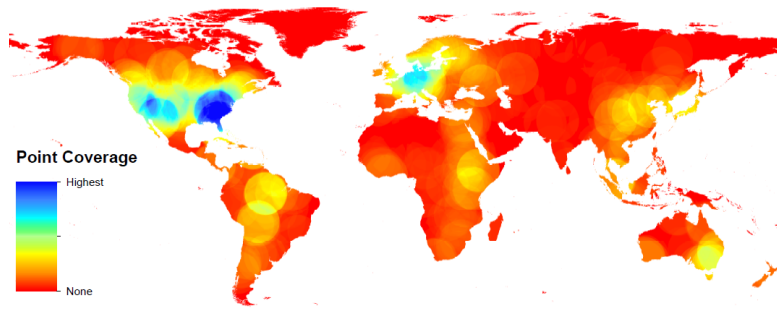
Figure 3. Global patterns of ET_OBS density for all LCs, as derived from the GETA 2.0 database.
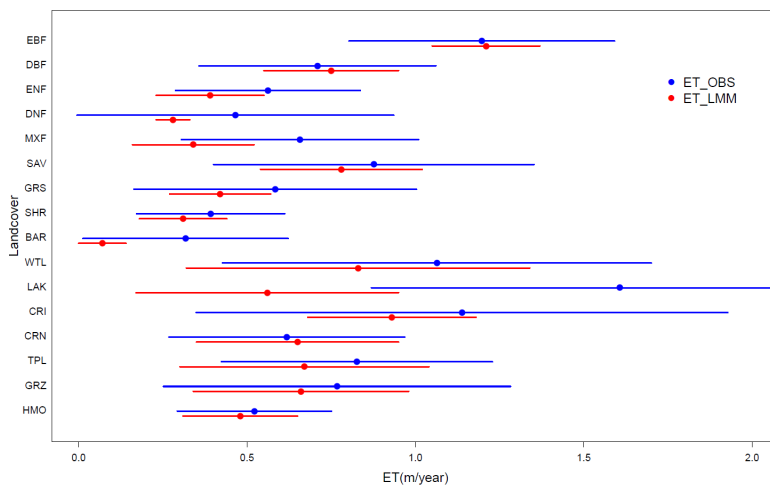
Figure 4. Mean (point) and standard deviation (line) of annual ET values by land-cover. ET_OBS represent point observations of ET from the GETA 2.0 database and ET_LMM represents ET of global fields generated by a statistical model from ET_OBS.
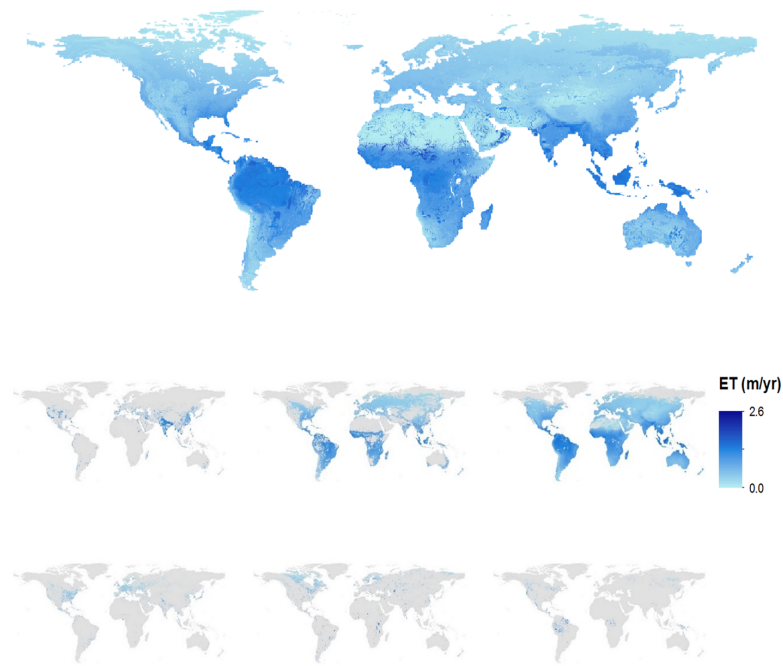
Figure 5. ET_LMM (m yr^{-1}) global annual actual ET projections. **(A)** Natural vegetation with wetland overlay. Grey indicates areas permanently covered by ice or large lakes. **(B)** ET_LMM for anthropogenic LCs. (i) Irrigated agriculture, (ii) non-irrigated agriculture, (iii) grazing land, (iv) urban lands, (v) lakes and reservoirs and (vi) tree plantations.
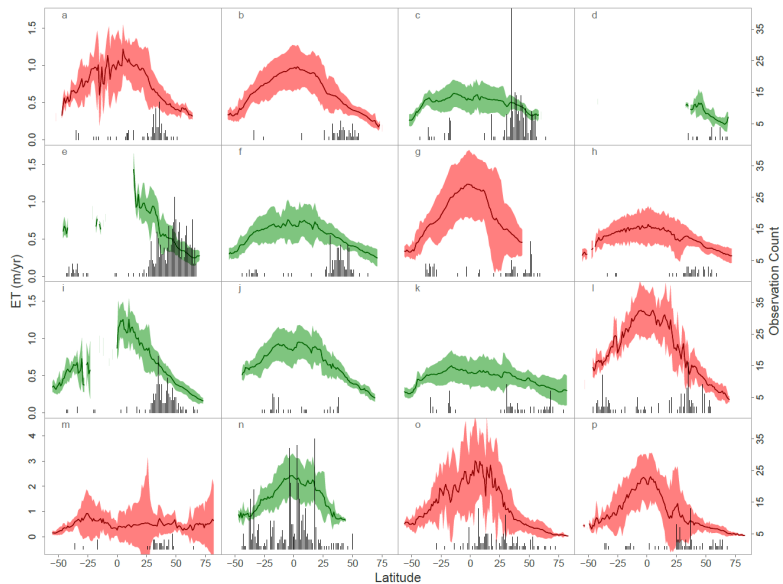
Figure 6. Zonal patterns of ET_LMM for LC types. The line represents the zonal mean, and the lighter area represents the 95 % confidence interval derived from bootstrapped data. Green represents natural vegetation (DBF, DNF, ENF, GRS, MXF, SAV, SHR, and EBF), and red represents anthropogenic vegetation (CRI, CRN, GRZ, HMO, TPL, BAR, LAK, (here classified as anthropogenic although it is a mix of natural lakes and anthropogenic reservoirs) and WTL) (see Table 1 for abbreviations). The black line plots represent the number of observations at that latitude. LC types are order alphabetically, except for the last row, which contains the LCs with the largest spread for ease of labelling (**a**: CRI, **b**: CRN, **c**: DBF, **d**: DNF, **e**: ENF, **f**: GRS, **g**: GRZ, **h**: HMO, **i**: MXF, **j**: SAV, **k**: SHR, **l**: TPL, **m**: BAR, **n**: EBF, **o**: LAK, **p**: WTL).
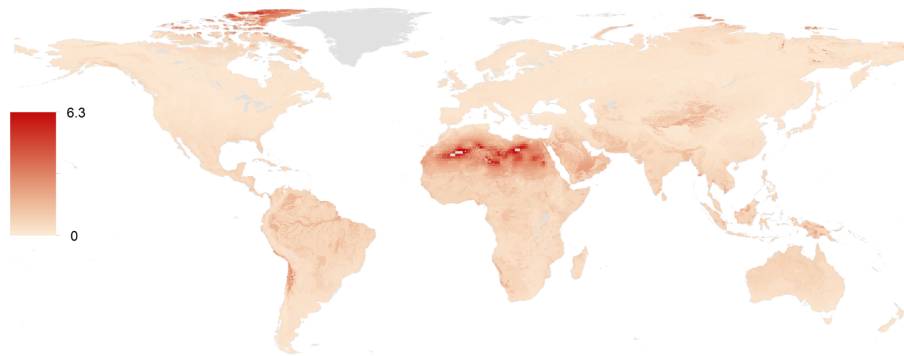
Figure 7. Uncertainty interval size for ET_LMM generated from 1000 bootstrapped datasets displayed with an equalized stretched histogram to emphasize the differences in location for potential vegetation.