

# Development and Validation of a Long-Term, Global, Terrestrial Sensible Heat Flux Dataset

AMANDA L. SIEMANN, NATHANIEL CHANEY, AND ERIC F. WOOD

*Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey*

(Manuscript received 30 October 2017, in final form 19 April 2018)

## ABSTRACT

Sensible heat flux is a turbulent flux driving interactions between the Earth's surface and the atmosphere, propelling local and regional climate. While turbulent fluxes are measured in situ, global scales require estimates at larger spatial scales, which can be made using remotely sensed satellite data. This study uses a first-order approximation to calculate the unconstrained hourly, terrestrial,  $0.5^\circ$ -resolution sensible heat flux using a land surface temperature consistent with the High Resolution Infrared Radiation Sounder (HIRS) retrievals, six reanalysis-based air temperature products, and a dataset of Zilitinkevich empirical constant  $C_{zil}$  values. This sensible heat flux dataset is constrained using the daily Bowen ratio and available energy, to produce nine constrained, daily products. All resulting global, terrestrial averages are within the uncertainty range of  $\pm 6.3 \text{ W m}^{-2}$  from the  $38.8 \text{ W m}^{-2}$  global annual average previously reported in the literature. The product constrained with the net radiation using the Moderate Resolution Infrared Spectroradiometer (MODIS) albedo and air temperature from the National Centers for Environmental Protection (NCEP) Climate Forecast System Reanalysis (CFSR) performs closest to the FLUXNET ground observations in the monthly analysis. These sensible heat flux estimates should be used for benchmarking global climate models at monthly or annual scales, and improvements should be made to the accuracy of input variables, particularly the temperature gradient,  $C_{zil}$  estimates, and the roughness length.

## 1. Introduction

The partitioning of available energy at the land surface produces the turbulent fluxes between the surface and the atmosphere, including sensible heat flux and latent heat flux, which transfer heat and water, respectively (Pipunic et al. 2008). While these fluxes are measured in situ using eddy covariance in tower networks (e.g., Oak Ridge National Laboratory 2016) and during field campaigns (e.g., the Boreal Ecosystem–Atmosphere Study; Sellers et al. 1997), satellite remote sensing provides more coverage for estimates at the global scale (Jiménez et al. 2011). Although the development of global, terrestrial estimates of latent heat flux has progressed (e.g., Jung et al. 2011; Wang and Dickinson 2012; Dolman et al. 2014; Miralles et al.

2016), the development of sensible heat flux global estimates encounters additional challenges, including the often poorly measured radiometric surface temperature and air temperature (Bastiaanssen et al. 1998), as well as parameterization of the aerodynamic resistance for heat transfer (Kustas et al. 1989).

Historically, several sensible heat flux algorithms based on remotely sensed data have used aircraft data for regional studies. Kustas et al. (1989) initially use thermal infrared data collected from an aircraft over Owens Valley, California, to apply the bulk transfer equation over heterogeneous land cover. Although Kustas et al. (1989) aim to increase accuracy in the aerodynamic resistance, which produced the large biases in the sensible heat flux with in situ data, they only test the new aerodynamic resistance formulation at two local sites and do not develop a method applicable across land-cover types and climate zones. Ogunjemiyo et al. (2003) also use meteorological and radiative observations collected from aircraft to estimate sensible heat flux using friction velocity, potential temperature, and Obukhov length. While Ogunjemiyo et al. (2003) find good agreement between the observations and model

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0732.s1>.

Corresponding author: Amanda L. Siemann, [siemann@princeton.edu](mailto:siemann@princeton.edu)

predictions made using a multiple regression model, the model requires phenological status and surface wetness of the land cover, as well as information on subsurface conditions, which is not often available.

Other algorithms use remotely sensed data, but utilize an energy balance model to estimate terrestrial sensible heat flux. Bastiaanssen et al. (1998) describe the Surface Energy Balance Algorithm for Land (SEBAL) to calculate sensible heat flux from hemispherical surface reflectance, surface temperature, and normalized difference vegetation index (NDVI), and Timmermans et al. (2007) compare regional results from SEBAL with those of the two-source energy balance (TSEB) model, which separates the radiometric temperature into soil and vegetation temperatures. While TSEB performs better than SEBAL in comparisons with flux tower observations at two sites, both models require proper selection of endmembers of NDVI or radiometric temperature, respectively, within a given scene, introducing additional sensitivities in the results to this choice and limiting the application and scale of these methods (Timmermans et al. 2007).

The desire to avoid the computational complications of numerical models and the reliance on surface meteorology motivate the use of the Atmospheric–Land Exchange Inversion (ALEXI) model, which relies on the remotely sensed observations of radiometric temperature at two times during each day, defining the daily amplitude of surface temperature. Mecikalski et al. (1999) produce regional estimates with ALEXI for the central contiguous United States (CONUS). Although Mecikalski et al. (1999) insist that the model is general enough to be applied to a range of surface types without calibrating regionally or making empirical adjustments per site, current implementations of the model require input temperature to closely resemble Geostationary Operational Environmental Satellite (GOES) temperatures over the CONUS. Other limitations of this model include the need for radiation data as an input to solve for the fluxes, the fact that the model has not been applied globally to our knowledge, and that the coverage of brightness temperatures and accuracy of these temperatures are limited by cloud contamination (Mecikalski et al. 1999). Norman et al. (2000) develop a model to simplify the implementation of the ALEXI model, but conclude that this dual-temperature-difference method is best applied using local input data, reinforcing the limitations of this model to the local or regional scale.

The transition from regional- to global-scale estimates of the energy budget, including sensible heat flux, has progressed in recent years by combining remote sensing data with various other data sources. Earlier studies (e.g., Trenberth et al. 2009) use reanalysis datasets, such

as the National Center for Atmospheric Research (NCAR) reanalysis 1 (NRA; Kalnay et al. 1996), the recent Japanese 25-yr Reanalysis (JRA; Onogi et al. 2007), and the second-generation 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA-40; Uppala et al. 2005), as the most comprehensive global estimates of the surface fluxes at that time. For a larger emphasis on observed data, Jung et al. (2011) utilize observed in situ and satellite data to assess global patterns of various fluxes, including sensible heat flux, for the vegetated land surface from 1982 to 2008. The observations used in Jung et al. (2011) are upscaled FLUXNET eddy-covariance observations, with the upscaling based on model tree ensembles (MTEs) trained on climate, land-use, meteorology, and remote sensing indices such as the fraction of absorbed photosynthetic active radiation (FAPAR) to create a 0.5°-resolution dataset, but restrict the study to monthly time series to improve computational efficiency and noise in the data. Jiménez et al. (2011) conduct a global comparison of 12 turbulent flux estimates using remotely sensed data, reanalysis data, and offline land surface models (LSMs) from 1993 to 1995, including the product formed in Jung et al. (2011), but several products do not directly model sensible heat flux, and the total range of  $39 \text{ W m}^{-2}$  is reported for the global average annual means based on 1994 in this study. Stephens et al. (2012) use the average of these 12 sensible heat estimates for terrestrial sensible heat flux and combine this with the SeaFlux estimates for the ocean to obtain global estimates from 2000 to 2010, but they note that no definitive measure of uncertainty in these flux estimates exists, and inconsistencies between products could produce the nonclosure of the energy balance.

Wild et al. (2015) and L'Ecuyer et al. (2015) present the most updated estimates of the global average terrestrial sensible heat flux over land and ocean, separately, in their energy budget estimates. Wild et al. (2015) use phase 5 of the Coupled Model Intercomparison Project (CMIP5) model estimates, which show a global average range spanning  $27 \text{ W m}^{-2}$ , but Wild et al. (2015) ultimately calculate the global average sensible heat flux estimate as a residual of the energy balance instead of explicitly calculating it. L'Ecuyer et al. (2015) average the estimates for sensible heat flux from the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004), the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011), and the Princeton evapotranspiration (ET) dataset using the Surface Energy Balance System (SEBS) model to calculate the sensible heat flux from 2003 to 2006

(Vinukollu et al. 2011). Although the time period of 2000–09 is longer than the time period of 2000–04 used in Wild et al. (2015), it only covers a fraction of the recent historical period in which remotely sensed data have been available. Additionally, neither of these estimates is closely based on observations for the sensible heat flux, except Vinukollu et al.'s (2011) estimate using SEBS, but this observation-based flux is only computed from 2003 to 2006.

Because these available global, terrestrial sensible heat flux data products are not based closely on observations, are estimated as a residual of other energy budget fluxes, have limited record lengths of 10 years or less, or have low temporal resolution at the monthly or annual scale, we develop our sensible heat flux data product to fill these gaps. In this study, we first develop and analyze a 0.5°, hourly, global, terrestrial sensible heat flux product from 1979 to 2009, which is aggregated (upscaled) to develop an improved, daily sensible heat flux data product from 1984 to 2007. We have greater confidence in the daily dataset because it is constrained by the available energy using the daily Bowen ratio. The development of this product is part of the World Climate Research Programme's (WCRP) Global Energy and Water Exchange (GEWEX) Data Assessment Panel (GDAP) activities focused on estimating the turbulent fluxes and understanding their variability. As GDAP strives for consistency between inputs and satellite data used across all datasets, we use a land surface temperature (LST) consistent with the High Resolution Infrared Radiation Sounder (HIRS) observations (Coccia et al. 2015; Siemann et al. 2016) that GDAP has selected for the updated Surface Radiation Budget version 4 (SRB-4; P. Stackhouse 2017, personal communication). We use six reanalysis-based air temperature products to form an ensemble of six sensible heat flux estimates, including optimized Zilitinkevich empirical constant values from 85 FLUXNET (Oak Ridge National Laboratory 2016) locations extended globally in an objective analysis, using climate and land-cover covariates (Chaney et al. 2016) in the aerodynamic resistance parameterization. The final, constrained sensible heat data product presented here is based on remotely sensed observations and calculated using a first-order bulk approximation; has a 24-yr record, which is longer than previous estimates, at daily temporal resolution; and extends over all land, except Antarctica and Greenland. After specifying all data and methods included in this study in section 2, we present annual averages of our products and the context of global annual averages within literature values in sections 3a and 3b for our unconstrained products and sections 3c and 3d for our constrained products. In section 3e, we examine seasonal variability, and in section 3f, we compare our products with reanalysis

products. Our product is validated against FLUXNET data in section 3g. These results are discussed in sections 4a–4g.

## 2. Data and methods

### a. First-order bulk approximation of sensible heat flux

In this study, we calculate the sensible heat flux  $H$  at the surface within the atmospheric boundary layer (ABL) according to Brutsaert (2005), in which Monin–Obukhov similarity theory can be used to parameterize the surface sensible heat flux as follows:

$$H = -\rho C_p C_h u_z \Delta\bar{\theta}, \quad (1)$$

where  $\rho$  is the air density,  $C_p$  is the specific heat capacity of air at constant pressure,  $C_h$  is the coefficient of heat transfer,  $u_z$  is the mean wind speed at a reference height, and  $\Delta\bar{\theta}$  is the difference between the potential temperatures at two different reference heights. By grouping the  $C_h$  and  $u_z$  into the aerodynamic resistance  $r_a$  as follows,

$$r_a = \frac{1}{C_h u_z}, \quad (2)$$

we parameterize the surface sensible heat flux using LST, air temperature, and aerodynamic resistance (Monteith 1973):

$$H = \frac{\rho C_p (T_s - T_a)}{r_a}. \quad (3)$$

Here,  $\rho$  ( $\text{kg m}^{-3}$ ) is calculated from the National Oceanic and Atmospheric Administration's (NOAA) Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) air pressure (Pa) at the 2-m air temperature (K);  $C_p$  is held constant at  $1004.6 \text{ J kg}^{-1} \text{ K}^{-1}$ ;  $T_s$  is the physical LST calculated using an HIRS-consistent LST (K) (Coccia et al. 2015) and the Moderate Resolution Imaging Spectroradiometer (MODIS) emissivity (because the HIRS-consistent LST assumes an emissivity of 1.0);  $T_a$  is the potential air temperature calculated using a 2-m air temperature (K); and  $r_a$  is calculated with units of seconds per meter. In Eq. (3), the  $(T_s - T_a)$  value is the vertical temperature gradient at the land surface between the physical LST and the air temperature at the land surface (calculated as the potential air temperature). Sections 2b–2d describe the datasets used for LST and  $T_a$ , as well as the method for calculating  $r_a$ .

### b. HIRS-consistent LST dataset

For the physical LST in all data products, we use an hourly, 0.5°-resolution LST product statistically consistent

TABLE 1. Sensible heat flux product acronyms with their respective description.

Sensible heat product	Description
Rs_Gr_CFSR	Remote sensing algorithm using CFSR temperature gradient
Rs_T_CFSR	Remote sensing algorithm using CFSR air temperature
Rs_T_MERRA	Remote sensing algorithm using MERRA air temperature
Rs_T_WZ_ERAInt	Remote sensing algorithm using Wang and Zeng air temperature based on ERA-Interim
Rs_T_WZ_NRA	Remote sensing algorithm using Wang and Zeng air temperature based on NRA
Rs_T_WZ_MERRA	Remote sensing algorithm using Wang and Zeng air temperature based on MERRA
Rs_Rnet_alM_Gr_CFSR	Remote sensing algorithm constrained using Rnet with MODIS albedo and CFSR temperature gradient
Rs_Rnet_alM_T_CFSR	Remote sensing algorithm constrained using Rnet with MODIS albedo and CFSR air temperature
Rs_Rnet_alM_T_MERRA	Remote sensing algorithm constrained using Rnet with MODIS albedo and MERRA air temperature
Rs_Rnet_alV_Gr_CFSR	Remote sensing algorithm constrained using Rnet with VIC albedo and CFSR temperature gradient
Rs_Rnet_alV_T_CFSR	Remote sensing algorithm constrained using Rnet with VIC albedo and CFSR air temperature
Rs_Rnet_alV_T_MERRA	Remote sensing algorithm constrained using Rnet with VIC albedo and MERRA air temperature
Rs_Rnet_alS_Gr_CFSR	Remote sensing algorithm constrained using Rnet with SRB albedo and CFSR temperature gradient
Rs_Rnet_alS_T_CFSR	Remote sensing algorithm constrained using Rnet with SRB albedo and CFSR air temperature
Rs_Rnet_alS_T_MERRA	Remote sensing algorithm constrained using Rnet with SRB albedo and MERRA air temperature
Obs_Tower	Tower observations
Rea_CFSR	CFS reanalysis product
Rea_MERRA	MERRA reanalysis product
Rs_Tforc_	Remote sensing algorithm using tower data and constrained as above with nine corresponding suffixes

with HIRS satellite retrievals. The retrievals were aggregated to  $0.5^\circ$ -resolution grids in hourly time steps and merged with the CFSR estimates at the same resolution using the model conditional processor (Coccia and Todini 2011; Todini 2008) Bayesian methodology proposed in Coccia et al. (2015). The final dataset extends continuously from 1979 to 2009 and includes the merged LST products corresponding to each satellite, averaged into one dataset using the linear opinion pool approach (Stone 1961; Clemen and Winkler 1999). We validated the final LST product against the Baseline Surface Radiation Network (BSRN)-based LST estimates (Siemann et al. 2016). Because the HIRS-consistent LST was formed assuming an emissivity of 1.0, the MODIS emissivity is utilized in this study to adjust the HIRS-consistent LST product into a physical LST. The MODIS emissivity is a monthly product at  $0.5^\circ$  resolution for the 14 MODIS University of Maryland (UMD) vegetation types (Ren et al. 2013).

### c. Reanalysis-based surface air temperature datasets

#### 1) DESCRIPTION OF DATASETS

We use six  $0.5^\circ$ -resolution, hourly, global 2-m air temperature datasets from 1979 to 2009 to calculate the first six sensible heat data products in Table 1, corresponding with each of the following air temperatures. We calculate the first air temperature at this resolution by subtracting the surface temperature gradient in CFSR output, computed from the CFSR 2-m air

temperature and the CFSR LST estimates, from the HIRS-consistent LST to back out an air temperature consistent with this CFSR temperature gradient. We use two of the other air temperatures directly from reanalysis products, including CFSR and MERRA (Rienecker et al. 2011) 2-m air temperature. We use the remaining three air temperatures developed by Wang and Zeng (2013). These three  $0.5^\circ$ -resolution, hourly land surface air temperature datasets are based on reanalysis products: the ECMWF interim reanalysis (ERA-Interim), the NRA, and MERRA, which were scaled to this consistent resolution in time and space and subsequently bias corrected by monthly mean minimum and maximum surface air temperatures using Climate Research Unit Time Series 3.10 monthly gridded data (Wang and Zeng 2013). See Table 1 for the acronyms for the sensible heat flux calculated using each air temperature product.

#### 2) COMPARISON WITH ISD STATION DATA

Because we calculate six sensible heat flux datasets using six different air temperatures, we validate each of the air temperature datasets against 13 545 stations in the NOAA National Centers for Environmental Information Integrated Surface Database (ISD) observations for air temperature (Smith et al. 2011) to determine how closely they compare with ground observations, globally. The overall average bias, root-mean-square error (RMSE), and Pearson correlation coefficient (CC) across all stations for all matching time periods of stations within

TABLE 2. Overall average Pearson CC, bias, and RMSE for all stations during 1979–2009 for each gridded air temperature dataset with the available ISD station data. The bolded font highlights the smallest bias, smallest RMSE, and largest CC, showing the product with the best performance using each statistic.

Air temperature dataset	CC	Bias (K)	RMSE (K)
CFSR	0.857	<b>−0.214</b>	3.920
HIRS-consistent using CFSR gradient	0.842	1.181	4.490
MERRA	0.855	0.539	4.083
Wang and Zeng based on ERA-Interim	<b>0.877</b>	−0.302	<b>3.620</b>
Wang and Zeng based on MERRA	0.871	−0.464	3.818
Wang and Zeng based on NRA	0.857	−0.224	3.995

1979–2009 are displayed in Table 2. The CFSR air temperature dataset has the smallest average bias, but the Wang and Zeng air temperature dataset, based on ERA-Interim, has the highest correlation coefficient and lowest RMSE with the ISD stations. Although these temperatures stand out, all six air temperatures fall within 2 K of the global average for the ISD station data, and all correlation coefficients are above 0.84. The HIRS-consistent air temperature using the CFSR temperature gradient and the HIRS-consistent LST shows the largest bias of 1.181 K, globally, as well as the smallest correlation coefficient of 0.842 and the largest RMSE of 4.490 K.

#### d. Aerodynamic resistance

We use the parameterization for aerodynamic resistance used in the offline Noah LSM version 3.4.1 as described in (Ek et al. 2003). This routine solves for the  $C_h$  using the following formulation based on Janjić (1994):

$$C_h = \frac{u_* k}{\left[ \ln \left( \frac{z + z_{0m}}{z_{0h}} \right) + \Psi_h \left( \frac{z + z_{0m}}{L} \right) - \Psi_h \left( \frac{z_{0h}}{L} \right) \right]}, \quad (4)$$

in which  $u_*$  is the friction velocity ( $\text{m s}^{-1}$ ),  $k$  is the von Kármán constant,  $z$  is the reference height (m), and  $z_{0h}$  and  $z_{0m}$  are the roughness lengths (m) for heat and momentum. The  $L$  in Eq. (4) is the Obukhov length that is described by the following equation (Monin and Obukhov 1954):

$$L = - \frac{u_*^3 \bar{\theta}_v}{kgw'\theta'_v}, \quad (5)$$

where  $g$  is gravitational acceleration ( $\text{m s}^{-2}$ ),  $\theta_v$  is the virtual potential temperature (K), and  $w'\theta'_v$  is the virtual potential temperature flux at the surface [ $\text{K (m}^2\text{s)}^{-1}$ ]. The heat stability correction function  $\Psi_h$  is based on Paulson (1970) for unstable boundary layer conditions and based on Holtslag and de Bruin (1988) for stable boundary layer conditions. In addition to applying Beljaars's (1995) correction for  $u_*$  to account for

instances of vanishing wind speed due to free convection, this parameterization also applies the following relationship between  $z_{0h}$  and  $z_{0m}$  to account for the difference between the radiative skin temperature (used as the LST) and the near-surface air temperature:

$$\frac{z_{0m}}{z_{0h}} = \exp(kC_{zil}\sqrt{\text{Re}^*}), \quad (6)$$

where  $C_{zil}$  is the Zilitinkevich empirical constant (Zilitinkevich 1995), and  $\text{Re}^*$  is the roughness Reynolds number. Instead of holding  $C_{zil}$  constant at 0.1, which is standard procedure in the Noah LSM following Chen et al. (1997), we use estimates calibrated at 85 FLUXNET eddy-covariance locations, which are extended globally at a 1-km spatial resolution using objective analysis based on climate and land-cover covariates [see Chaney et al. (2016) for further details.] Additional inputs include the specific humidity, which is calculated using the air temperature calculated from applying the CFSR temperature gradient to the HIRS-consistent LST to be consistent with this LST and CFSR data, and the evapotranspiration to calculate a correction on the 2-m specific humidity to convert it to the specific humidity at the surface, for which we use the hourly CFSR evapotranspiration output to remain consistent with the other CFSR inputs.

#### e. Vegetation weighted average, green-vegetation fraction, and postprocessing

We use the roughness length for momentum  $z_{0m}$  from the UMD land-cover-type dataset (Defries et al. 2000) for 14 land-cover types upscaled from  $\sim 1$  km to  $0.5^\circ$  resolution [see Hansen et al. (2000) for detailed description of each type]. To calculate the total sensible heat flux for each grid cell, we run the algorithm for each land-cover type present in that grid cell. We compute a weighted average total sensible heat flux value for each cell based on the fractional green-vegetation coverage over each land-cover type within the cell [derived from Gutman and Ignatov (1998)], which is then added to the sensible heat flux calculated for the bare soil fraction.



For evergreen broadleaf forest and bare soil, the solver for aerodynamic resistance may not converge with 50 iterations, or the  $C_h$  may be large, causing a very large sensible heat flux. We check for unreasonable sensible heat fluxes over  $1000.0 \text{ W m}^{-2}$ , under  $-1000.0 \text{ W m}^{-2}$  (the absolute value is well above the maximum incoming solar radiation), or for which the aerodynamic resistance does not converge within 50 iterations. In these cases, we use an average  $C_h$  for the respective vegetation type. If an average  $C_h$  is not available for the grid cell, we use an overall global average  $C_h$ . If the sensible heat flux weighted average remains greater than or equal to  $1000.0 \text{ W m}^{-2}$  or less than or equal to  $-1000.0 \text{ W m}^{-2}$ , then the estimate is made undefined.

#### f. Data used to constrain sensible heat flux estimates

As seen later in sections 3a and 3b, several of the global sensible heat flux averages overestimate literature values as well as reanalysis values. These results prompt us to constrain our sensible heat flux estimates while still preserving the proportion of energy partitioned into sensible heat flux when using our algorithm based on remotely sensed data. We base this constraint on the energy budget at the land surface, defined by Liang et al. (2010) as

$$R_n - G = H + L, \quad (7)$$

in which  $R_n$  is net radiation,  $G$  is ground heat flux, and  $L$  is latent heat flux, all in watts per square meter. At the daily scale, we solve for the constrained  $H$  ( $H_c$ ) as follows:

$$H_c = \frac{R_n - G}{1 + \frac{1}{\text{abs}(B)}}, \quad (8)$$

in which  $B$  is the daily Bowen ratio, defined by Bowen (1926) as

$$B = \frac{H}{L}. \quad (9)$$

We use the absolute value of the Bowen ratio so that the  $H_c$  is never larger than the available energy because we are trying to constrain the sensible heat flux by the available energy. Using these equations, we create nine daily constrained sensible heat flux products from 1984 to 2007 (due to the limits of the net radiation data record) using the daily Bowen ratio and the daily available energy. These nine products are described in rows 7–15 in Table 1 as those constrained with net radiation using one of three albedo datasets and one of three air temperature datasets. Sections 2f(1)–2f(3) describe the

datasets used for the net radiation, ground heat flux, and latent heat flux.

#### 1) NET RADIATION

We calculate the net radiation products at daily,  $0.5^\circ$  resolution using the longwave (LW) and shortwave (SW) downward components from the SRB-3 products (Stackhouse et al. 2011). For the SW upward radiation, we use the SRB-3 product and apply one of three different albedo products. For one set of radiation products, and therefore one set of constrained sensible heat flux products, we use the SW upward radiation as is, which we label as the SRB albedo product. For another set of radiation products, we apply the MODIS albedo as a monthly climatology, derived from Liang and Liu (2012), Qu et al. (2014), N. Liu et al. (2013), and Q. Liu et al. (2013). For the last of the three sets of radiation products, we apply the albedo averaged from hourly to daily scale from the Variable Infiltration Capacity model, forced with CFSR land. For all products, we derive the LW upward radiation from the daily averaged HIRS-consistent LST and the land-cover-averaged MODIS emissivity (Ren et al. 2013) using the Stefan–Boltzmann law (Campbell and Norman 1998).

#### 2) GROUND HEAT FLUX

We calculate a 3-hourly average consistent ground heat flux product using the formulation based on Sellers (1965):

$$G(0, t) = \Delta T_0 \sqrt{\omega C \lambda} \sin\left(\omega t + \frac{\pi}{4}\right), \quad (10)$$

where  $\Delta T_0$  is the amplitude of the surface temperature wave during each day [calculated as  $\frac{1}{2}(T_{\max} - T_{\min})$ ],  $\omega$  is the frequency of the wave (with a 24-h period) ( $\text{s}^{-1}$ ), and  $t$  is the time of day (s), and aggregate it up to daily resolution. We calculate the thermal inertia  $\Gamma$ , equal to  $\sqrt{C \lambda}$  where  $C$  is the volumetric heat capacity [ $\text{J (m}^3 \text{K)}^{-1}$ ] and  $\lambda$  is the soil thermal conductivity [ $\text{W (m K)}^{-1}$ ], using the following formulation in Murray and Verhoef (2007):

$$\Gamma = K_e (\Gamma_* - \Gamma_0) + \Gamma_0, \quad (11)$$

where  $K_e$  is the Kersten number,  $\Gamma_*$  is the saturated thermal inertia, and  $\Gamma_0$  is the dry thermal inertia.

Murray and Verhoef (2007) calculate the Kersten number using the following equation derived from Lu et al. (2007):

$$K_e = \exp[\gamma'(1 - S_r^{-\delta'})], \quad (12)$$

where  $\gamma'$  is the approximated soil texture parameter,  $\delta'$  is the approximated shape parameter, and  $S_r$  is the

TABLE 3. The categories of soil wetness taken from Murray and Verhoef (2007) used for the calculation of the thermal inertia for the ground heat flux. The percentiles are of soil moisture values taken from the ECDF of all the annual average soil moisture, globally. The values are the soil moisture values that correspond to each percentile threshold. The  $S_r$  is the fractional water filled porosity assigned from the center of the range given for each category in Murray and Verhoef (2007).

Soil	Percentiles	Values ( $\text{m}^3 \text{m}^{-3}$ )	$S_r$
Dry	<20th	<0.1274	0.05
Dry-moist	$\geq 20\text{th}$ and <40th	0.1274–0.242 399	0.17
Moist	$\geq 40\text{th}$ and $\leq 60\text{th}$	0.2424–0.3794	0.375
Moist-wet	>60th and $\leq 80\text{th}$	0.397 401–0.5361	0.625
Wet	>80th	>0.5361	0.875

fractional water-filled porosity. Murray and Verhoef (2007) base their approximations for  $\gamma'$  and  $\delta'$  on fitting Eq. (11) to the curve relating thermal inertia to the  $S_r$ , making the  $\delta' = 1.78$  and the  $\gamma' = 2.0/\delta'$  for sand fraction over 0.8; the  $\delta' = 0.93$  and the  $\gamma' = 1.5/\delta'$  for sand fraction under 0.4; and the  $\delta' = 3.84$  and the  $\gamma' = 4.0/\delta'$  for sand fraction of 0.4–0.8. The sand fraction is taken from a lookup table in Murray and Verhoef (2007) based on the soil texture. We use the SoilGrids global dataset (Hengl et al. 2014) aggregated up to 50 km for the soil texture. Murray and Verhoef (2007) approximate the  $S_r$  using the soil moisture categories listed in Table 3. We use the thresholds of soil moisture based on percentiles of soil moisture taken from the global empirical cumulative distribution function (ECDF) of all annual average soil moisture values ( $\text{m}^3 \text{m}^{-3}$ ) from 1984 to 2007 for these categories of soil moisture to obtain the corresponding  $S_r$ . We classify each cell based on the weekly average soil moisture calculated from the European Space Agency (ESA) Climate Change Initiative (CCI) passive soil moisture daily product (Liu et al. 2012; Wagner et al. 2012), upscaled from  $0.25^\circ$  to  $0.5^\circ$  spatial resolution. Additionally,  $\Gamma_*$  and  $\Gamma_0$  are calculated as follows, according to Murray and Verhoef (2007):

$$\Gamma_* = 788.2\theta_*^{-1.29}, \quad \text{and} \quad (13)$$

$$\Gamma_0 = -1062.4\theta_* + 1010.8, \quad (14)$$

where  $\theta_*$  is the saturated soil moisture content taken from the lookup table using soil texture in Murray and Verhoef (2007).

### 3) LATENT HEAT FLUX

We compute the daily latent heat flux using the Penman–Monteith model used by Sheffield et al. (2010) as specified in the following equation:

$$L = \frac{R_n + \frac{\rho c_p \text{VPD}}{r_a}}{\Delta + \gamma \left(1 + \frac{r_s}{r_a}\right)}, \quad (15)$$

where VPD is the vapor pressure deficit (Pa),  $\Delta$  is the slope of the saturated vapor pressure versus the air temperature ( $\text{Pa K}^{-1}$ ), and  $r_s$  is the canopy resistance ( $\text{s m}^{-1}$ ). All inputs are consistent with the sensible heat flux products using the same HIRS-consistent LST, net radiation described above, meteorology for the input data from CFSR, air temperatures consistent with the sensible heat flux products, and the aerodynamic resistance, all averaged to the daily scale to calculate the latent heat flux at  $0.5^\circ$  resolution.

### g. FLUXNET validation dataset

Sensible heat flux is one of the fluxes measured with the eddy covariance method at a global network of over 650 meteorological sites (Baldocchi 2008). For this study, we use the La Thuile dataset, which is a summary database of a subset of 253 stations at 30-min resolution with data from 1996 to 2007 where available (Oak Ridge National Laboratory 2016). We use this sensible heat flux data to compare with our constrained sensible heat flux estimates. We also use the meteorology to force our algorithm to assess how well our algorithm compares with observed sensible heat flux at the collocated tower grid using these data collected at the tower. When we aggregate this dataset up to daily resolution and filter for daily energy budget closure within 10%, we find 12 stations with over 5 years of data that pass the quality controls, 33 stations with over 3 years of data, 106 stations with over 1 year of data, 154 stations with over 6 months of data, and 201 stations with over 90 days of data. These stations are color coded in Fig. 1 by the minimum threshold length of quality-controlled data that each station passes. Supplemental Fig. S1 shows the difference in monthly statistics of  $\text{Rs\_Rnet\_alM\_T\_CFSR}$  with tower observations by different record-length thresholds. We choose to use the 6-month threshold for our analysis because 6 months' minimum includes multiple data points in our monthly analysis (often more than six, as quality-controlled data are not required to be consecutive) and multiple seasons. Additionally, 41 of the stations that pass the 6-month quality control have observed data at the towers for all necessary input variables for our algorithm (air temperature, LST derived from LW radiation, specific humidity, and wind speed), so we run our algorithm at these stations with tower input data for comparison.

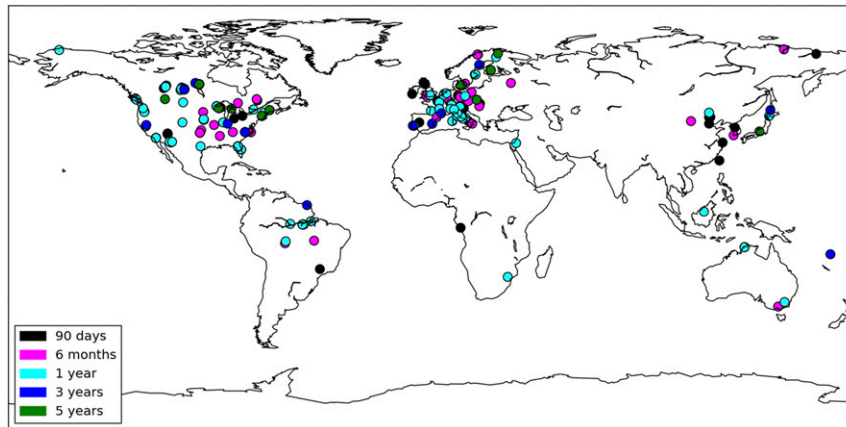


FIG. 1. Map of the FLUXNET locations that pass five minimum quality-control thresholds of data record length. All analysis shown in section 3 includes all stations shown here having over 6 months, 1 year, 3 years, and 5 years of data, as the minimum threshold chosen is 6 months. Fig. S1 includes all stations in this map in the 90-day-threshold box plot.

### 3. Results—Assessment of sensible heat datasets

#### a. Annual averages of sensible heat flux datasets

Figure 2 displays the annual average sensible heat flux calculated at  $0.5^\circ$  resolution from 1979 to 2009 for each of the six sensible heat flux estimates: namely,  $R_s\_Gr\_CFSR$  (Fig. 2a),  $R_s\_T\_CFSR$  (Fig. 2b),  $R_s\_T\_MERRA$  (Fig. 2c),  $R_s\_T\_WZ\_ERAInt$  (Fig. 2d),  $R_s\_T\_WZ\_NRA$  (Fig. 2e), and  $R_s\_T\_WZ\_MERRA$  (Fig. 2f). Five of the six products (in Figs. 2b–f) include negative sensible heat fluxes in the northern high latitudes, although products in Fig. 2b and Figs. 2d–f include an eastern region of these high latitudes with positive sensible heat fluxes. Additionally, several products include negative sensible heat fluxes in areas of high altitude, such as the Andes Mountains, high-elevation regions of China, and parts of the Tibetan Plateau. All sensible heat fluxes are under  $50 \text{ W m}^{-2}$  in the forests of the Congo, and in some parts under  $20 \text{ W m}^{-2}$ , and under  $70 \text{ W m}^{-2}$  in the forests of the Amazon and portions of Indonesia. In Figs. 2b–f, the average sensible heat fluxes of the Congo and portions of Indonesia are negative in some regions. Figures 2b–f display the highest average sensible heat fluxes in Australia, the Horn of Africa, the Sahara, much of the Middle East, and areas of central Asia, such as portions of India and China. Figure 2a displays the highest average sensible heat fluxes in the Sahara, parts of the Middle East, and regions of central Asia and Australia, although these higher sensible heat fluxes are at least  $20 \text{ W m}^{-2}$  less than the same regions in Figs. 2b–f, with the largest differences in Australia. Average sensible heat fluxes in Figs. 2d–f are at least  $80 \text{ W m}^{-2}$ , reaching up to  $125 \text{ W m}^{-2}$  in portions of the Southwest CONUS, Great Plains, upper Midwest CONUS, and Canadian

Plains. When comparing these six products with one another, Fig. 2a displays the lowest overall averages for sensible heat flux globally.

#### b. Annual global averages compared with literature estimates

We are interested not only in the average spatial patterns of these terrestrial products, but also in how well these average estimates compare with other global terrestrial averages from the literature. When comparing products with literature estimates and maintaining consistency among the comparisons, we need to appropriately account for estimates in Greenland and Antarctica, which are excluded in our six product estimates due to lack of forcing data in those regions, but which are included in global terrestrial estimates from Wild et al. (2015) and L'Ecuyer et al. (2015). We use the annual average for Antarctica and Greenland,  $-33.2 \text{ W m}^{-2}$ , from the CFSR sensible heat flux to calculate a weighted average based on area that is consistent with the terrestrial global averages from the literature. The left side of Table 4 displays the annual global averages for each product, excluding Antarctica and Greenland or including Antarctica and Greenland. The bottom two rows show the averages reported in the literature for comparison. Consistent with Fig. 2, the lowest of our product averages is  $R_s\_Gr\_CFSR$  at  $34.1 \text{ W m}^{-2}$  with Antarctica and Greenland, which nearly matches the CFSR sensible heat average of  $34.7 \text{ W m}^{-2}$ . The second-lowest average is  $R_s\_T\_MERRA$ , which is  $8.1 \text{ W m}^{-2}$  above the largest unconstrained literature average of  $38.8 \text{ W m}^{-2}$  (L'Ecuyer et al. 2015). All remaining products fall between 56.9 and  $61.5 \text{ W m}^{-2}$ . Because five of the six mean annual global averages are at least  $18.0 \text{ W m}^{-2}$  above the largest



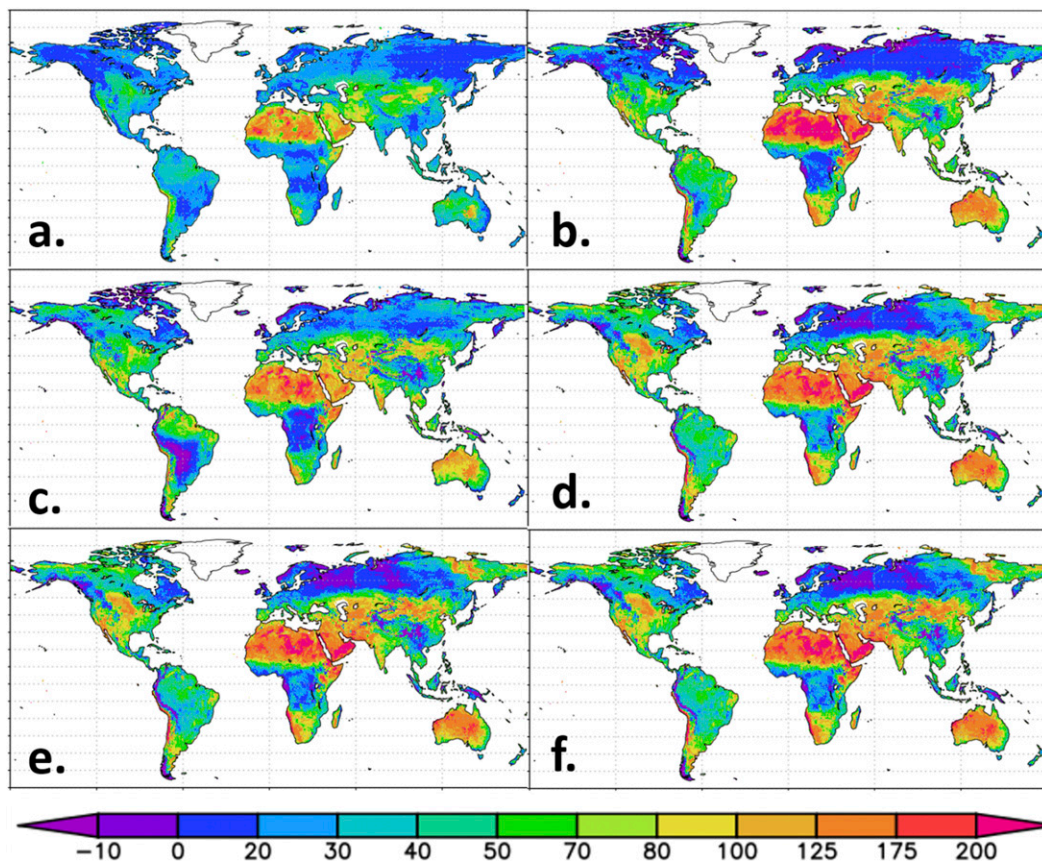


FIG. 2. The annual average sensible heat flux ( $\text{W m}^{-2}$ ) 1979–2009 for our six products based on using the (a) CFSR temperature gradient, (b) CFSR air temperature, (c) MERRA air temperature, (d) Wang and Zeng air temperature based on ERA-Interim, (e) Wang and Zeng air temperature based on NRA, and (f) Wang and Zeng air temperature based on MERRA.

literature value, we constrain our products following the methodology described in section 2f, producing nine daily, constrained sensible heat flux products.

#### c. Annual averages of constrained sensible heat flux datasets

Figure 3 displays the annual average sensible heat flux calculated at  $0.5^\circ$  resolution from 1984 to 2007 for each of the nine constrained sensible heat flux estimates: namely,  $\text{Rs\_Rnet\_alM\_Gr\_CFSR}$  (Fig. 3a),  $\text{Rs\_Rnet\_alM\_T\_CFSR}$  (Fig. 3b),  $\text{Rs\_Rnet\_alM\_T\_MERRA}$  (Fig. 3c),  $\text{Rs\_Rnet\_alV\_Gr\_CFSR}$  (Fig. 3d),  $\text{Rs\_Rnet\_alV\_T\_CFSR}$  (Fig. 3e),  $\text{Rs\_Rnet\_alV\_T\_MERRA}$  (Fig. 3f),  $\text{Rs\_Rnet\_alS\_G\_CFSR}$  (Fig. 3g),  $\text{Rs\_Rnet\_alS\_T\_CFSR}$  (Fig. 3h), and  $\text{Rs\_Rnet\_alS\_T\_MERRA}$  (Fig. 3i). While all nine products have average sensible heat fluxes under  $20 \text{ W m}^{-2}$  in the northern high latitudes, none of the products have negative average sensible heat fluxes, unlike the unconstrained annual averages in Fig. 2. All sensible heat fluxes are around

$10 \text{ W m}^{-2}$  less than the unconstrained estimates in the forests of the Congo at under  $30\text{--}40 \text{ W m}^{-2}$ , except for the body of water in the west, which is up to  $60\text{--}70 \text{ W m}^{-2}$ . The estimates in the Amazon using the CFSR temperature gradient and the MERRA air temperature are all under  $50 \text{ W m}^{-2}$ , while those using the CFSR temperature are below  $60 \text{ W m}^{-2}$ , except for the rivers (up to  $90 \text{ W m}^{-2}$ ) and the easternmost region of South America (up to  $60\text{--}70 \text{ W m}^{-2}$ ). All panels display the highest average sensible heat fluxes in the Sahara, the Horn of Africa, southern Africa, the Arabian Peninsula, portions of central Asia, and portions of Australia. Figures 3a–c and 3g–i have some averages over  $100 \text{ W m}^{-2}$  in the Sahara and the Arabian Peninsula, while Figs. 3d–f show almost all averages in the Sahara and Arabian Peninsula over  $100 \text{ W m}^{-2}$  (often up to 130 or  $150 \text{ W m}^{-2}$ ). The Horn of Africa displays averages over  $100 \text{ W m}^{-2}$  in all products. The averages in parts of central Asia, the Andes, western South Africa, and Australia are all between  $70$  and  $100 \text{ W m}^{-2}$ , with a few

TABLE 4. Updated table of overall mean annual global averages ( $\text{W m}^{-2}$ ) for our six sensible heat products and our nine constrained products, excluding Antarctica and Greenland (excl. A&G) or including Antarctica and Greenland (incl. A&G) for comparison with literature values. We use the CFSR sensible heat flux average for Antarctica and Greenland of  $-33.2 \text{ W m}^{-2}$ .

SH model	excl. A&G ( $\text{W m}^{-2}$ )	incl. A&G ( $\text{W m}^{-2}$ )	Constrained SH	excl. A&G ( $\text{W m}^{-2}$ )	incl. A&G ( $\text{W m}^{-2}$ )
Rs_Gr_CFSR	42.3	34.1	Rs_Rnet_alM_Gr_CFSR	44.7	36.2
Rs_T_CFSR	67.9	56.9	Rs_Rnet_alM_T_CFSR	51.6	42.4
Rs_T_MERRA	55.9	46.2	Rs_Rnet_alM_T_MERRA	48.9	40.0
RS_T_WZ_MERRA	73.0	61.5			
Rs_T_WZ_NRA	68.9	57.9	Rs_Rnet_alV_Gr_CFSR	46.6	38.0
Rs_T_WZ_ERAIInt	70.8	59.5	Rs_Rnet_alV_T_CFSR	53.3	43.9
Other Products			Rs_Rnet_alV_T_MERRA	50.7	41.6
Rea_CFSR	43.0	34.7			
Rea_MERRA	42.2	34.0	Rs_Rnet_alS_Gr_CFSR	46.4	37.7
Wild et al. (2015)		32.0	Rs_Rnet_alS_T_CFSR	53.2	43.9
L'Ecuyer et al. (2015) constrained		38.1	Rs_Rnet_alS_T_MERRA	50.6	41.5

regions above  $100 \text{ W m}^{-2}$  and up to about  $120 \text{ W m}^{-2}$ , depending on the product. Some regions show drastically reduced annual averages, compared with the unconstrained estimates in Fig. 2, including the Southwest CONUS and the Great Plains under  $80 \text{ W m}^{-2}$ , with most areas under  $50\text{--}60 \text{ W m}^{-2}$ ; the upper Midwest CONUS under  $50 \text{ W m}^{-2}$ ; and the Canadian Plains under  $50\text{--}60 \text{ W m}^{-2}$ . When comparing the products using different air temperatures, the products using the CFSR temperature gradient (left) have the lowest annual averages for sensible heat flux, globally, while averages

using the CFSR air temperature (center) are around  $10 \text{ W m}^{-2}$  larger in many regions, and the averages using the MERRA air temperature (right) are in between.

*d. Global averages of constrained sensible heat flux datasets compared with literature estimates*

Table 4 displays the annual global averages for the constrained products, as well as the literature estimates, reanalysis estimates, and the unconstrained estimates. The constrained averages using the CFSR air temperature are the highest sensible heat flux averages among

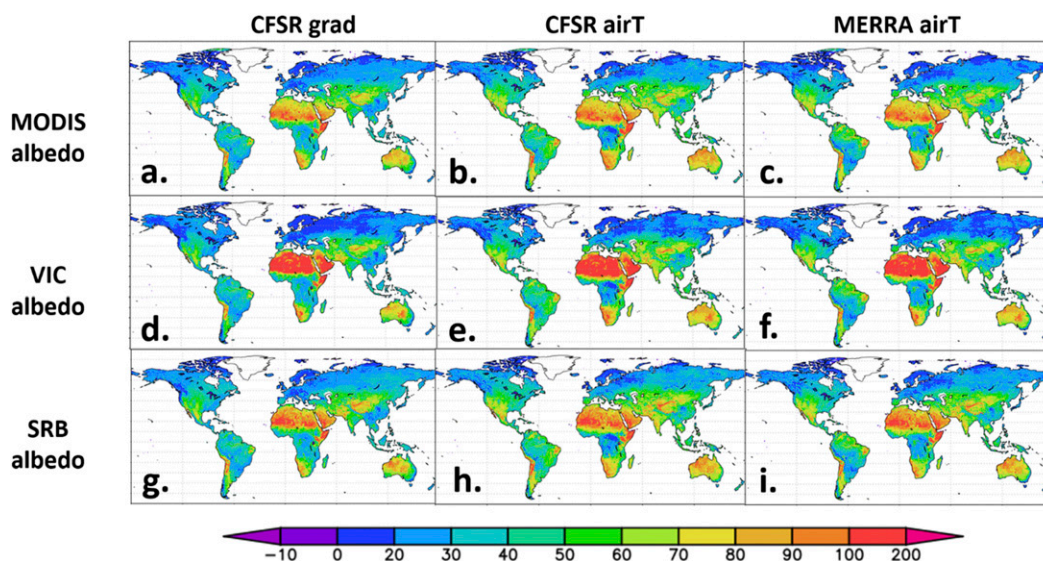


FIG. 3. The constrained annual average sensible heat flux ( $\text{W m}^{-2}$ ) 1984–2007 for our nine products using the (a) CFSR temperature gradient and the net radiation with the MODIS albedo, (b) CFSR air temperature and the net radiation with the MODIS albedo, and (c) MERRA air temperature and the net radiation with the MODIS albedo. (d)–(f) using the same air temperatures as (a)–(c), but using the VIC albedo, and (g)–(i) using the same air temperatures as (a)–(c), but using the SRB albedo. Note the different color scale between 50 and 200 from the color scale in Fig. 2 to show the spatial variations in the data with the appropriate resolution for the constrained data range. Note:  $1 \text{ W m}^{-2} = 31.5471 \text{ MJ (m}^2 \text{ yr)}^{-1}$ .



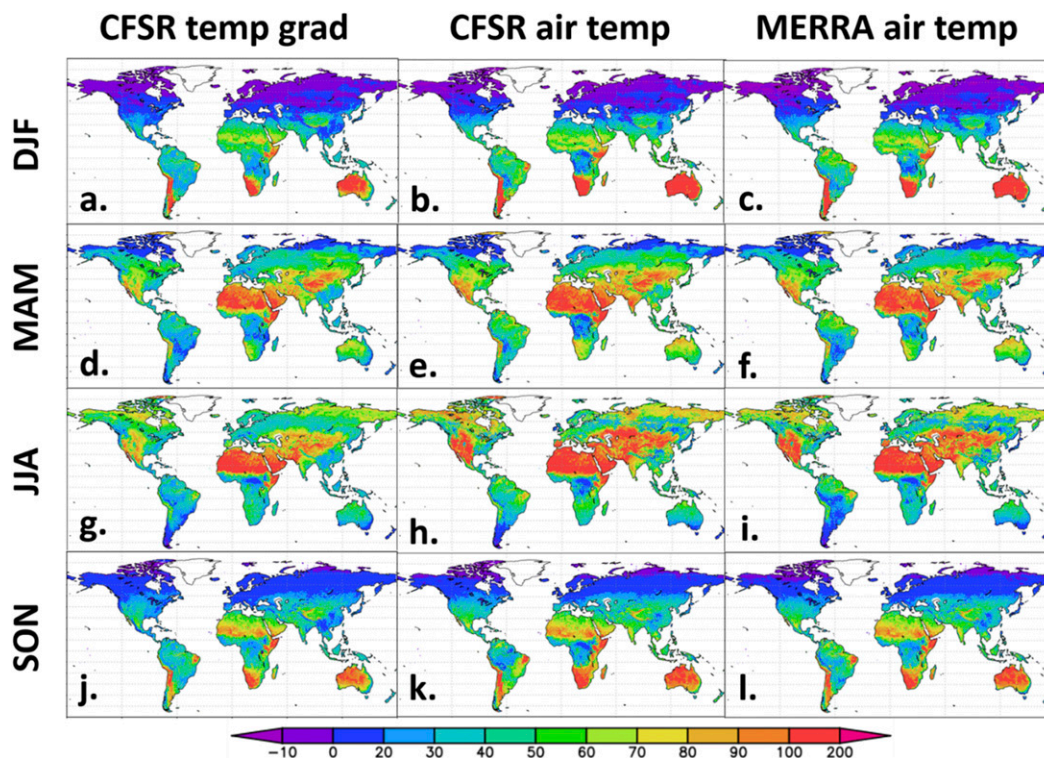


FIG. 4. The seasonal average sensible heat flux ( $\text{W m}^{-2}$ ) 1984–2007 for the constrained sensible heat flux products calculated with the (left) CFSR temperature gradient, (center) CFSR air temperature, and (right) MERRA air temperature, all using the net radiation with the MODIS albedo for the 3-month seasons of DJF, MAM, JJA, and SON (top to bottom). Note:  $1 \text{ W m}^{-2} = 31.5471 \text{ MJ (m}^2 \text{ yr)}^{-1}$ .

the constrained products, while the constrained averages using the CFSR temperature gradient are among the lowest sensible heat flux averages among the constrained products. The products using the MODIS albedo have the lowest overall global averages for each air temperature among the constrained products, which are all within or under  $4 \text{ W m}^{-2}$  above the range of literature averages. The largest global average among the constrained products is only  $5.1 \text{ W m}^{-2}$  above the largest literature average, showing the drastic reduction in the averages, compared with those of the unconstrained products. Resulting from the improved range of averages relative to the literature values, we will only focus on the constrained sensible heat flux data products in sections 3e–3g.

#### e. Seasonal variability in sensible heat flux datasets

Seasonal averages of the sensible heat flux provide insight into the dynamics of the seasonal cycle within our estimates. Figure 4 displays the four seasonal averages for our sensible heat flux products calculated with  $\text{Rs\_Rnet\_alM\_Gr\_CFSR}$ ,  $\text{Rs\_Rnet\_alM\_T\_CFSR}$ , and  $\text{Rs\_Rnet\_alM\_T\_MERRA}$ . The largest positive sensible heat fluxes in the Southern Hemisphere are in the September–November (SON) and December–February

(DJF) seasons, while the largest positive sensible heat fluxes in the Northern Hemisphere are in the March–May (MAM) and June–August (JJA) seasons in all datasets. The most negative sensible heat flux estimates are in parts of the northern high latitudes in DJF and SON. When comparing seasonal patterns among products, all products show similar seasonal spatial patterns and seasonal changes.

#### f. Grid-by-grid comparison of annual averages with reanalysis data

To compare our sensible heat estimates to other available estimates on a grid-by-grid basis, we calculate overall Pearson correlation coefficient and average bias of our sensible heat flux products  $\text{Rs\_Rnet\_alM\_Gr\_CFSR}$ ,  $\text{Rs\_Rnet\_alM\_T\_CFSR}$ , and  $\text{Rs\_Rnet\_alM\_T\_MERRA}$  with the reanalysis sensible heat fluxes, displayed in Fig. 5. When comparing our estimates with both reanalysis products, estimates in the Sahara and Arabian Peninsula are between  $5$  and  $50 \text{ W m}^{-2}$  higher than either reanalysis estimate and occasionally over  $50 \text{ W m}^{-2}$  higher. In comparisons with  $\text{Rea\_MERRA}$  in Figs. 5j–l, our estimates have positive biases between  $15$  and  $50 \text{ W m}^{-2}$  in northern South America, Indonesia,

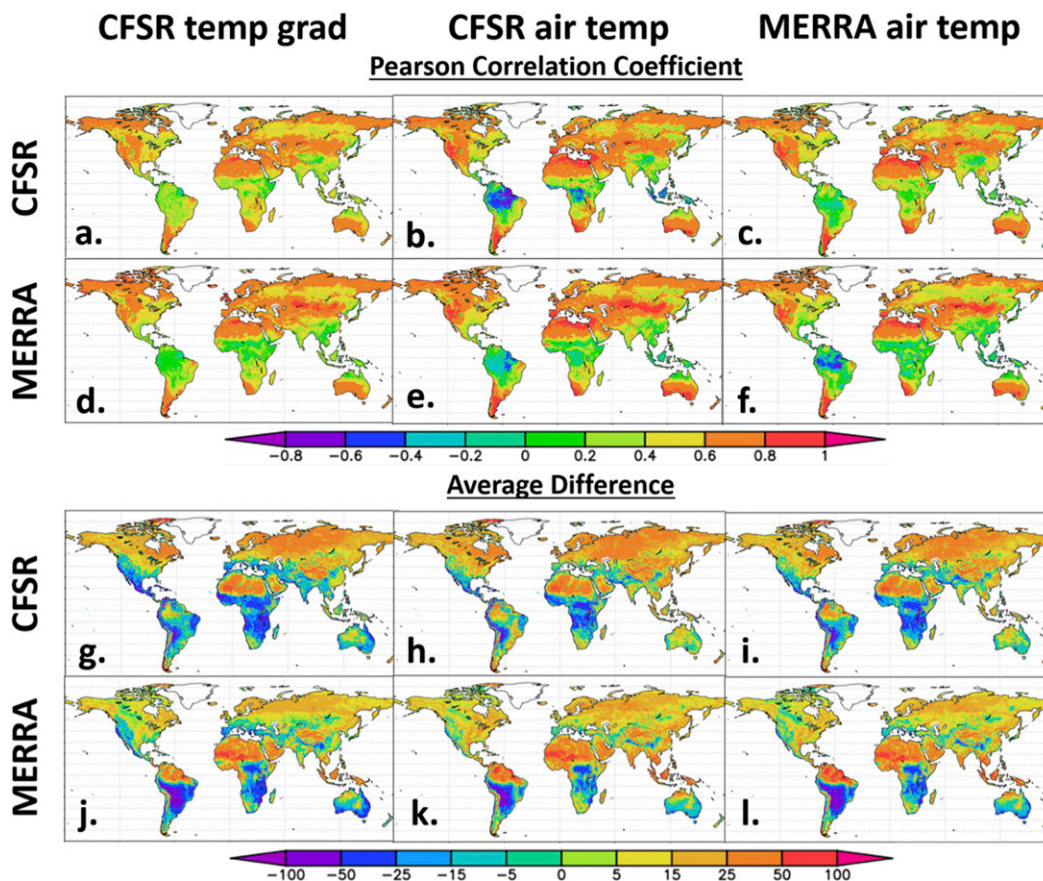


FIG. 5. Pearson CCs and average bias ( $\text{W m}^{-2}$ ) of our sensible heat flux product using the (left) CFSR temperature gradient, (center) CFSR air temperature, and (right) MERRA air temperature, and all constrained using the net radiation with the MODIS albedo with the CFSR and MERRA reanalysis sensible heat fluxes. Note:  $1 \text{ W m}^{-2} = 31.5471 \text{ MJ (m}^2 \text{ yr)}^{-1}$ .

southern Asia, and portions of central Asia. Average biases reach  $15\text{--}50 \text{ W m}^{-2}$  in the northern latitudes of Eurasia and eastern North America, when compared with Rea\_CFSR in Figs. 5g–i.

Although our products overestimate fluxes relative to reanalysis in some regions, our products underestimate fluxes relative to reanalysis in central Africa as well as in central South America by  $5\text{--}50 \text{ W m}^{-2}$ , producing average biases of  $-50$  to  $-100 \text{ W m}^{-2}$  in a small area of central South America when compared with Rea\_CFSR (Figs. 5g–i), but in the majority of central South America when compared with Rea\_MERRA (Figs. 5j–l). In the Northern Hemisphere, Figs. 5g and 5j show negative biases in regions of North America, western Europe, and portions of the Middle East, southern Asia, and Australia for the products using the CFSR temperature gradient, when compared with both reanalysis datasets, while these regions often have average biases between  $-5$  and  $5 \text{ W m}^{-2}$  (and at most within  $\pm 15 \text{ W m}^{-2}$ ) for products using the other two air temperature datasets.

Although the average differences may be larger than  $15 \text{ W m}^{-2}$ , suggesting poor performance, the correlations for the corresponding grids are above  $0.6\text{--}0.8$  in the Sahara, the Arabian Peninsula, and the high northern latitudes. The more intuitive situation of both statistics showing poor performance occurs in central Africa, the majority of South America, and Indonesia, with low or negative correlations (under  $0.4$  and often near zero) and large average differences. Generally, the highest correlation coefficients are seen in all products at latitudes above  $30^\circ\text{N}$  in North America, above  $20^\circ\text{N}$  and under  $20^\circ\text{S}$  in Africa, above  $20^\circ\text{N}$  in the Arabian Peninsula, above about  $30^\circ\text{--}40^\circ\text{N}$  in Asia, under  $30^\circ\text{S}$  in South America, and under  $20^\circ\text{S}$  in Australia.

#### g. Validation of sensible heat flux against observations using FLUXNET tower data

Observations from our nine constrained products based on remote sensing are compared with tower observations (Obs\_Tower) at the 154 FLUXNET towers

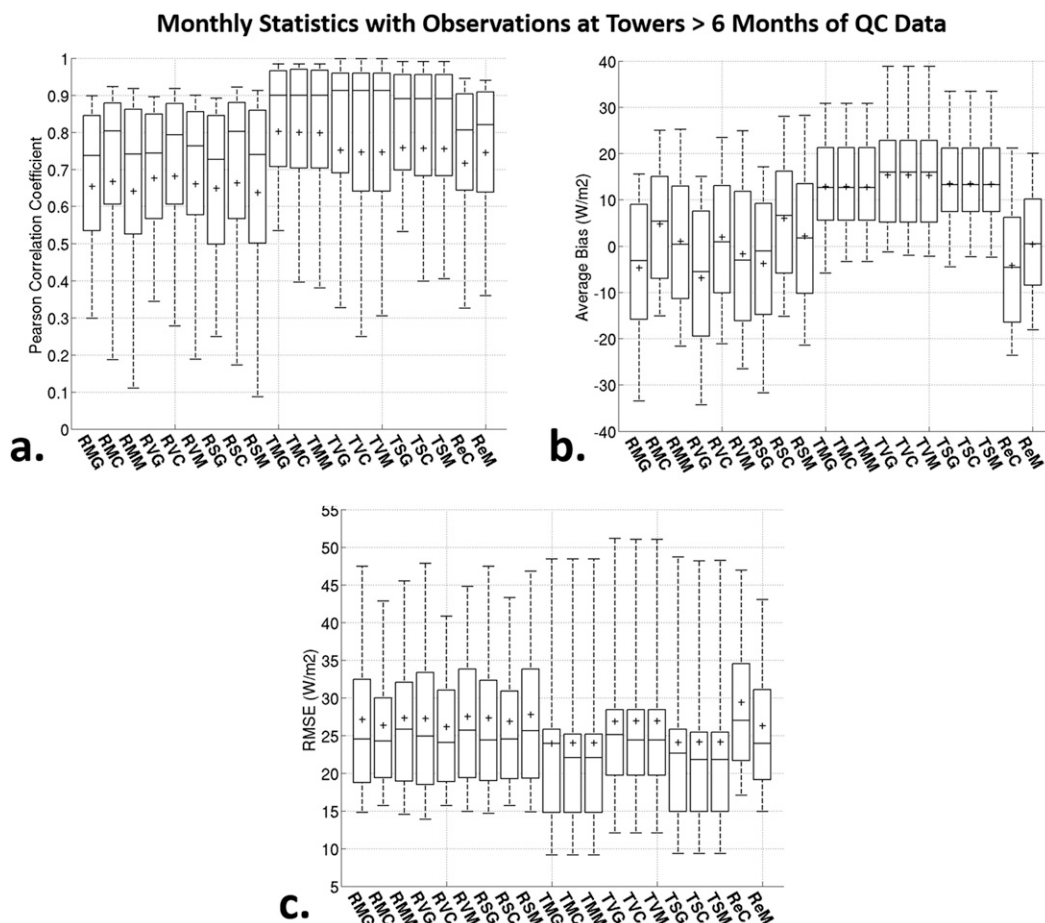


FIG. 6. Box plots displaying the (a) Pearson CC, (b) average bias ( $\text{W m}^{-2}$ ), and (c) RMSE ( $\text{W m}^{-2}$ ) for the nine constrained sensible heat flux products, nine constrained tower-forced sensible heat flux products, and two reanalysis sensible heat flux products with the FLUXNET observations using monthly averages at all towers that have at least 6 months of data passing quality control. The horizontal bars inside the boxes are the medians, the + signs are the means, the box edges are the 25th and 75th percentiles, and the whiskers show 9% and 91% of the data. Abbreviations for the products are done using the following format: for three capital letter abbreviations, the first letter is R for remote sensing or T for tower; the second letter is the M, V, or S for the MODIS, VIC, or SRB albedos; and the last letter G, C, or M is for the CFSR gradient, CFSR, or MERRA for the air temperature. In the last two labels, the first two letters, Re, are for reanalysis, and the following letter C or M is for CFSR or MERRA.

that have over 6 months of data passing quality control. We also compare the Rea\_CFSR and Rea\_MERRA datasets with Obs\_Tower at the full 154 stations, and the datasets forced with the tower data (Rs\_Tforc\_) with the Obs\_Tower at the 41 stations with sufficient input data at these stations. The box plots of the distributions of Pearson correlation coefficients, average biases, and root-mean-square errors with Obs\_Tower at all 154 stations (or 41 stations, in the case of the tower-forced datasets) using the monthly averages are displayed in Fig. 6. In Fig. S2, statistics are also calculated at the daily scale by removing seasonality using a 31-day moving window (averaged from 15 days on each side of the current day).

In Fig. 6a, in all cases, our algorithm run with tower inputs has a higher median (above 0.9) and higher mean (0.75–0.8) Pearson correlation coefficient than the corresponding algorithm run with remotely sensed inputs and constrained using the same latent heat flux, ground heat flux, and net radiation. While all the median correlation coefficients for the tower-forced algorithms (above 0.9) and the mean correlation coefficients for the products constrained using the net radiation with the MODIS albedo (at approximately 0.8) are larger than those of the reanalysis datasets, the tower-forced datasets using the VIC or SRB albedos have similar means (between 0.7 and 0.8) as the reanalysis means. Among the datasets using remotely sensed data, the median



correlation coefficient of around 0.8, and mean correlation coefficient of around 0.65 for the Rs\_Rnet\_alM\_T\_CFSR, Rs\_Rnet\_alV\_T\_CFSR, and Rs\_Rnet\_alS\_T\_CFSR are the highest. Although the mean correlation coefficients are higher for the reanalysis datasets (between 0.7 and 0.8), the median for Rea\_CFSR is similar (around 0.8) to the medians for Rs\_Rnet\_alM\_T\_CFSR, Rs\_Rnet\_alV\_T\_CFSR, and Rs\_Rnet\_alS\_T\_CFSR.

While median and mean correlation coefficients for tower-forced datasets are generally higher than or equal to the correlation coefficients for the datasets forced with remote sensing data and the reanalysis datasets, the median and mean average biases (Fig. 6b) are all larger (between 10 and 20  $\text{W m}^{-2}$ ) than those forced with remote sensing data and the reanalysis datasets (all within  $\pm 10 \text{ W m}^{-2}$ ). Among the datasets forced with remotely sensed data, those constrained using the MODIS albedo have median and mean average biases all within approximately  $\pm 5 \text{ W m}^{-2}$ , which are similar to the medians and means for both reanalysis datasets. Among the tower-forced datasets, those constrained with the MODIS albedo and SRB albedo have the lower median and mean average biases near  $10 \text{ W m}^{-2}$ , but the 25th percentile is closer to zero for the datasets constrained with the MODIS albedo. Rs\_Rnet\_alM\_G\_CFSR, Rs\_Rnet\_alM\_T\_MERRA, Rs\_Rnet\_alV\_T\_CFSR, Rs\_Rnet\_alV\_T\_MERRA, Rs\_Rnet\_alS\_G\_CFSR, and Rs\_Rnet\_alS\_T\_MERRA produce the smallest biases with Obs\_Tower of the datasets forced with remotely sensed data.

The mean and median RMSEs for all datasets are within 20–30  $\text{W m}^{-2}$ . The tower-forced datasets have the lowest median and mean RMSEs when compared with both reanalysis datasets and the datasets using remotely sensed inputs. Rs\_Rnet\_alM\_T\_CFSR and Rs\_Rnet\_alV\_T\_CFSR have the lowest mean and median RMSEs among the datasets using remotely sensed inputs, which are approximately equal to those of Rea\_MERRA and lower than those of Rea\_CFSR, while the tower-forced datasets using the VIC albedo have the highest mean and median RMSEs among the tower-forced datasets.

Figure 7 shows the Pearson correlation coefficients, average biases, and RMSEs among our datasets using remotely sensed data and the reanalysis datasets against Obs\_Tower based on vegetation type, region, and Köppen climate class (Peel et al. 2007). In all correlation coefficient panels, the mean and median correlation coefficients of the reanalysis datasets are higher than the corresponding mean and median for our datasets. When divided by vegetation type (Fig. 7a), both the reanalysis datasets and our datasets show highest median correlation coefficients at the wetland and water sites (over 0.9), evergreen forest sites (over 0.8), and shrubland and savanna sites (over

0.75). The highest means above 0.6 occur in the wetland and water sites, shrubland and savanna sites, and cropland and bare soil sites for our datasets and above 0.75 in the wetland and water sites, shrubland and savanna sites, and mixed and evergreen forest sites for the reanalysis datasets. In Fig. 7b, our datasets show the lowest median and mean correlation coefficients (below 0.5) in Africa and South America, while the reanalysis datasets show the lowest correlation coefficients in South America, but the highest mean and median correlation coefficients in Africa. Among climate classes, all datasets have the lowest mean and median correlation coefficients in the tropical class, while all datasets show the highest median and correlation coefficients in the arid class.

In Fig. 7d, the mean and median average biases are largest for all datasets in the grassland (over  $10 \text{ W m}^{-2}$ ), wetland and water (under  $-10 \text{ W m}^{-2}$ ), and cropland and bare soil (over  $5 \text{ W m}^{-2}$ ) vegetation types. The largest mean and median average biases are seen in Africa and Australia for both our datasets and the reanalysis datasets, although the magnitude of the biases are opposite, with our datasets showing negative biases (around  $-30 \text{ W m}^{-2}$ ) and reanalyses showing positive biases (around  $20 \text{ W m}^{-2}$ ). The mean and median biases in South America are above  $10 \text{ W m}^{-2}$  for our datasets, while the reanalysis mean and median biases are around  $0 \text{ W m}^{-2}$ . In Fig. 7f, the largest biases for the reanalysis datasets are seen in the arid towers, while the polar, arid, and tropical stations show similar large median biases (around  $10 \text{ W m}^{-2}$ ), with the largest mean bias (above  $15 \text{ W m}^{-2}$ ) at the polar towers.

When considering RMSEs, Fig. 7d shows the largest mean and median RMSEs at or above  $30 \text{ W m}^{-2}$  for our datasets at shrubland and savanna sites and all datasets at the grassland sites. The reanalyses have the lowest mean and median RMSEs at the wetland and water sites. In Fig. 7h, African and Australian sites have the largest mean RMSEs for our datasets (above  $35 \text{ W m}^{-2}$ ), while the reanalysis datasets have the largest mean and median RMSEs for the Australian sites and the South American sites (at or above 40 and  $35 \text{ W m}^{-2}$ , respectively). Figure 7i shows the largest median RMSEs in the arid and polar categories for our datasets above  $30 \text{ W m}^{-2}$ , while the means for all categories except the cold sites are above  $30 \text{ W m}^{-2}$ . The median and mean RMSEs are largest for the reanalysis datasets at the tropical and temperate sites, but all are within  $25\text{--}35 \text{ W m}^{-2}$  among these datasets.

## 4. Discussion

### a. Annual averages of sensible heat flux datasets

When examining the spatial patterns in the overall annual averages of our six sensible heat flux estimates in

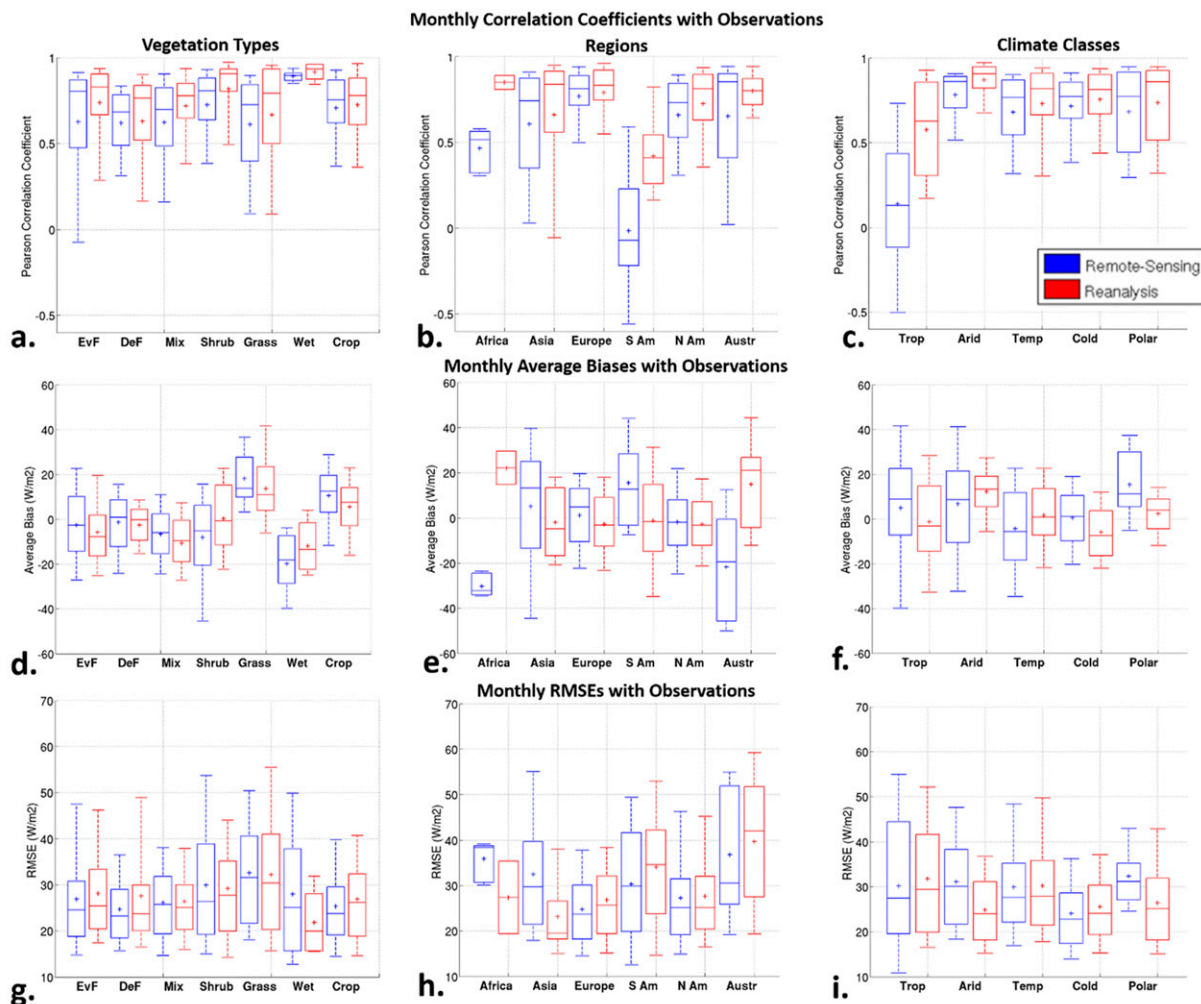


FIG. 7. Box plots displaying the (a)–(c) Pearson CC, (d)–(f) average bias ( $\text{W m}^{-2}$ ), and (g)–(i) RMSE ( $\text{W m}^{-2}$ ) for the nine constrained sensible heat flux products shown together in the blue-colored bar and the two reanalysis sensible heat flux products together in the red-colored bar with the FLUXNET observations using monthly averages at all towers that have at least 6 months of data passing quality control. The median, mean, box size, and dotted lines are the same style as Fig. 6. (a),(d),(g) Statistics by vegetation types including Evergreen Forest (EvF), Deciduous Forest (DeF), Mixed Forest (Mix), Shrubland and Savannas (Shrub), Grassland (Grass), Wetlands and Water (Wet), and Cropland and Bare Soil (Crop). (b),(e),(h) Statistics by region including Africa, Asia, Europe, South and Central America (S Am), North America (N Am), and Australia (Austr). (c),(f),(i) Statistics by Köppen climate class including Tropical (Trop), Arid, Temperate (Temp), Cold, and Polar.

Fig. 2, we find negative fluxes in regions in northern high latitudes (i.e., parts of Eurasia and North America) and in areas of high altitude (i.e., high-altitude areas of China and Andes Mountains) where snow may cover the land surface, causing a colder surface temperature than the air above it that manifests itself in a negative temperature gradient in Eq. (3). As noted in Serreze et al. (2000), snow-covered regions are primarily located north of  $50^{\circ}\text{N}$ , and our northern high-latitude negative sensible heat flux estimates are consistent in location with these snow-covered regions. While snow may explain negative sensible heat fluxes, the negative fluxes

are seen in patches in the regions above  $50^{\circ}\text{N}$ , and all regions are not covered by negative sensible heat averages consistently across all products. Small errors in temperature datasets may cause these sensible heat flux estimates of unexpected sign, likely producing these varying negative and positive patches in the high northern latitudes that vary between products, as well as small pockets of negative sensible heat flux in the Congo and Indonesia in our estimates in Figs. 2b–f and negative sensible heat fluxes in areas of central South America in Fig. 7c. Average biases as small as  $0.2\text{--}1.2\text{ K}$ , seen for the air temperature datasets when compared with ISD

station data, or average biases in the LST dataset of 0.27–3.44 K at individual BSRN stations (Siemann et al. 2016), can produce temperature gradients of the wrong sign, compared with the true gradient. Additionally, none of the temperature datasets have been well validated against ground observations in these locations. For example, no BSRN stations in these regions were used in Siemann et al. (2016) for the HIRS-consistent LST, and ISD station coverage is sparse in these locations. Thus, we cannot fully quantify the errors in the temperature datasets, which could be larger than those stated above. HIRS retrievals used to form the HIRS-consistent LST product are also sparse in the tropics, contributing to potential errors in the temperature gradient in the Congo (Siemann et al. 2016).

Although no studies of comparable record or methodology of data product formation have been conducted that provide flawless comparison, Jung et al. (2011) and Jiménez et al. (2011) provide some observation-based products for comparison in spatial patterns of sensible heat flux, globally. As mentioned in section 1, Jung et al. (2011) assessed global patterns of various fluxes, including sensible heat flux, from 1982 to 2008 using upscaled FLUXNET eddy-covariance observations produced with model tree ensembles, creating a 0.5°-resolution dataset [note: the conversion to the units in Jung et al. (2011) is  $1 \text{ W m}^{-2} = 31.5471 \text{ MJ (m}^2 \text{ yr)}^{-1}$ ]. The mean annual sensible heat flux patterns in this dataset include the highest sensible heat fluxes in Australia, the Middle East, South Africa, the Horn of Africa, the Sahel, and the Southwest CONUS, which are all consistent regions of high relative sensible heat flux in our products.

Consistent patterns of sensible heat fluxes in some regions relative to others do not always mean consistent magnitudes of the fluxes. Sensible heat fluxes over 50–70  $\text{W m}^{-2}$  in Fig. 2a and over 80  $\text{W m}^{-2}$  in Fig. 2c are comparable to the at least 60–80  $\text{W m}^{-2}$  seen in Jung et al. (2011) in northern Australia, but our estimates in Figs. 2b and 2d–f of over 100–125  $\text{W m}^{-2}$  overshoot the estimates of Jung et al. (2011) in this region. Magnitudes in the Sahel, South Africa, and the Middle East of about 60–80  $\text{W m}^{-2}$  are comparable to the estimates of at least 60–80  $\text{W m}^{-2}$  in the products in Figs. 2a and 2c, but our estimates for all other products are at least 20  $\text{W m}^{-2}$  higher than these estimates of Jung et al. (2011). In the cases above, the uncertainty in these regions in the sensible heat flux of 25–35  $\text{W m}^{-2}$  in Fig. 2 of Siemann et al. (2018) accounts for the higher sensible heat fluxes of Figs. 2b and 2d–f in these regions, as Fig. 3b in Siemann et al. (2018) shows the majority of the sensible heat flux uncertainty in these areas is attributed to temperature gradient uncertainty.

Areas of largest positive mismatch include the Midwest CONUS, Great Plains, Canadian Plains, and central Asia, in which estimates in Jung et al. (2011) can be as low as 20  $\text{W m}^{-2}$  in the Midwest CONUS and the Canadian Plains or 30–45  $\text{W m}^{-2}$  in the Great Plains and central Asia, compared with estimates in our products as high as 50–70  $\text{W m}^{-2}$  in Figs. 2a–c or above 80 or 125  $\text{W m}^{-2}$  in Figs. 2d–f. Siemann et al. (2018) describe large sensitivities to changes in  $C_{\text{zil}}$ , the temperature gradient, and the roughness length in the Great Plains, Canadian Plains, and central Asia. While any errors in all three inputs could influence this mismatch, even small errors in one input could push the sensible heat out of the range of data from the literature in areas with these high sensitivities. Siemann et al. (2018) attribute the largest fraction of uncertainty in the Great Plains, Canadian Plains, and central Asia to the temperature gradient and in the Midwest CONUS to  $C_{\text{zil}}$ , indicating these factors likely contribute most to the overestimation. Additionally, of all the factors, the  $C_{\text{zil}}$  cannot be measured, so it is possible that the optimization may not have produced a reliable value in these regions or for these land-cover types.

#### *b. Annual global averages compared with estimates from literature*

Considering the global, terrestrial averages of sensible heat flux from our six products in the left side of Table 4, the two products in closest range to the estimates from Wild et al. (2015) and L'Ecuyer et al. (2015) (32 and 38.8  $\text{W m}^{-2}$ , respectively) are Rs\_Gr\_CFSR and Rs\_T\_MERRA. Because the CFSR temperature gradient is consistent with the output temperature gradient used in the runs for the CFSR sensible heat flux, we expect the near equality of the global average of 34.07  $\text{W m}^{-2}$  of our product with the global average of Rea\_CFSR of 34.70  $\text{W m}^{-2}$ . These global averages are not identical for various reasons; for example, in the Noah LSM used for CFSR, the  $C_{\text{zil}}$  value was held constant at 0.1 for all times and all grid cells, in contrast to the  $C_{\text{zil}}$  dataset of optimized values, which vary substantially in space, and the vegetation-related parameter of roughness length for momentum may not be consistent between the CFSR Noah LSM runs and our climatology from the UMD dataset. The MODIS emissivity climatology used to calculate the physical LST from the HIRS-consistent LST may also be different from the emissivity used in CFSR, causing some differences in the true temperature gradient. In the case of Rs\_T\_MERRA, differences between the HIRS-consistent LST and the MERRA LST (see section 4c) could cause the disparity in temperature gradient, which would lead to the disparity in sensible heat flux in addition to the  $C_{\text{zil}}$  and roughness

length differences. The meteorological inputs for wind speed, humidity, and air pressure are also different between CFSR and MERRA, so these differences could altogether account for the global average difference of  $R_s\_T\_MERRA$  with  $Rea\_MERRA$ .

As seen in the left side of Table 4, the differences of the remaining four products with the two literature estimates range between 18.1 and  $29.5 \text{ W m}^{-2}$ , globally. These large differences, in addition to the different patches of negative sensible heat fluxes between products and the large differences noted in section 4a in several regions between our products and the literature estimates in Jung et al. (2011), prompt us to constrain our estimates using the methodology in section 2f.

### c. Annual averages of constrained sensible heat flux datasets

Many of the inconsistencies between the six unconstrained sensible heat flux products and literature values are improved in the nine constrained products. The northern high-latitude regions that are often negative in the unconstrained products are between 0 and  $20 \text{ W m}^{-2}$  (particularly in the western and high northern regions of Eurasia) and up to  $30 \text{ W m}^{-2}$  in the constrained products, which are consistent with the estimates in Jung et al. (2011) of around  $0\text{--}16 \text{ W m}^{-2}$  in the western and northernmost regions of Eurasia and up to around  $30 \text{ W m}^{-2}$  in the remaining regions of these northern high latitudes. Our constrained products share the highest average sensible heat fluxes in the Horn of Africa, western South Africa, the Arabian Peninsula, portions of central Asia, northern areas of Australia, and parts of the Sahara [vegetated fragments that are shown in Jung et al. (2011)] with the highest average sensible heat fluxes in Jung et al. (2011). The scale on the figure in Jung et al. (2011) reaches at or above around  $80 \text{ W m}^{-2}$  in some regions within those mentioned above, but our products likely overestimate by around  $20 \text{ W m}^{-2}$  in the Arabian Peninsula and the Horn of Africa [most of the Sahara is not included in Jung et al. (2011)]. The estimates in Australia, central Asia, the Andes, and South Africa are similar in range ( $70\text{--}100 \text{ W m}^{-2}$ ) to the approximately  $55\text{--}80+ \text{ W m}^{-2}$  in these regions in Jung et al. (2011). Additionally, regions with drastic reductions in annual averages between the unconstrained products and the constrained products, such as the Southwest CONUS, the Great Plains, the upper Midwest CONUS, and the Canadian Plains, match more closely with estimates in Jung et al. (2011) at  $60\text{--}80 \text{ W m}^{-2}$  in the Southwest CONUS and the Great Plains, and only  $20 \text{ W m}^{-2}$  above Jung et al. (2011) in the upper Midwest CONUS and the Canadian Plains. These remaining mismatches of around  $20 \text{ W m}^{-2}$  in these

regions can be attributed to the sensitivities discussed in section 4a.

Spatial patterns of lower sensible heat fluxes in Jung et al. (2011) in the Amazon and the Congo, including the higher sensible heat fluxes in the easternmost area of South America, are consistent in location with our nine constrained products. The eastern portion of South America shows different fluxes than the surrounding regions of the Amazon in all our constrained products [and in Jung et al. (2011)] because the land cover is woody savanna, savanna, and grassland, instead of forest. While the spatial patterns of lower sensible heat fluxes in the Congo and Amazon are consistent with greater partitioning of available energy into the latent heat flux in these regions, the smaller sensible heat fluxes in the Congo relative to those in the Amazon in our products have the opposite relationship in Jung et al. (2011). These differences between the fluxes in the Congo and Amazon could result from the sparse coverage of satellite data used in the HIRS-consistent LST or lack of validation in the tropics noted in section 4a, as well as the largest uncertainty for the sensible heat flux in the Amazon shown in Siemann et al. (2018). Regarding other features in these regions, rivers or areas classified as water land cover show higher sensible heat fluxes in our products because the roughness is drastically different for this land-cover type from the surrounding forests, and the latent heat flux we use to constrain the sensible heat fluxes is drastically lower than the surrounding forest pixels (not shown).

### d. Global averages of constrained sensible heat flux datasets compared with literature estimates

Seen both in section 3c and in the right side of Table 4, the largest sensible heat flux global averages using the CFSR air temperature and the smallest sensible heat flux global averages using the CFSR temperature gradient are caused by the relative magnitudes of the air temperatures shown in Table 2. Relative to the ISD station data, the CFSR temperature gradient has the most positive bias of the three air temperature datasets used in the constrained products, while CFSR air temperature has a negative bias. The larger air temperature would produce a smaller temperature gradient with the same HIRS-consistent LST according to Eq. (3), as seen in our products.

Although much closer in range to the literature averages, these global averages differ by up to 4.2 or  $5.1 \text{ W m}^{-2}$ , depending on which products are compared with which literature value of the two. In addition to the differences in the temperature gradient between our datasets and the literature data, the products in the literature may have inconsistent masks, different time spans from which the annual averages were calculated [only 2000–04



in Wild et al. (2015) and 2000–09 in L'Ecuyer et al. (2015)], and different annual averages for Greenland and Antarctica from the average we choose to use from CFSR. Although only the Rs\_Rnet\_alM\_G\_CFSR product average is within the uncertainty range around the value in Wild et al. (2015) of  $25\text{--}36\text{ W m}^{-2}$ , all product averages are within the uncertainty range of  $\pm 6.3\text{ W m}^{-2}$  of the  $38.8\text{ W m}^{-2}$  average in L'Ecuyer et al. (2015).

*e. Seasonal variability sensible heat flux datasets*

Considering seasonal differences in Rs\_Rnet\_alM\_Gr\_CFSR, Rs\_Rnet\_alM\_T\_CFSR, and Rs\_Rnet\_alM\_T\_MERRA, the largest positive sensible heat fluxes in the Northern Hemisphere in MAM and JJA and in the Southern Hemisphere in SON and DJF (see Fig. 4) are expected due to the larger available energy during those summer months in the respective hemispheres. Larger available energy means more energy will go into the turbulent fluxes, so positive increases in sensible heat flux during those seasons are likely, relative to winter seasons. Although energy is partitioned into sensible heat flux in the winter, if the LST is cooler than the air temperature, this gradient produces negative sensible heat fluxes, while warmer LSTs in the summer have more potential to produce positive sensible heat fluxes. The negative sensible heat fluxes in Fig. 4 are only present in the northern high latitudes in SON and DJF. This location is consistent with the snow-covered regions above  $50^\circ\text{N}$  (Serreze et al. 2000) mentioned in section 4a, and the spatial trend of more negative sensible heat fluxes farther north is consistent with Ye et al. (1998), who use historical records of weather station data across Eurasia from 1881 to 1985 to show the spatial trend of increased average snow depth toward higher latitudes. Additionally, according to Dye (2002), the average week of first-observed snow cover between 1972 and 2000 in the northern high latitudes of Eurasia is the week of 30 September at earliest above  $60^\circ\text{N}$  and in the northern high latitudes of North America can be as early as the week of 9 September. These dates are consistent with the start of the negative sensible heat fluxes in these regions during SON in Figs. 4j–l. If Fig. 4 used products constrained with the net radiation using the SRB or VIC albedos, the patterns of the fluxes would be similar using the SRB albedo, but the fluxes would increase by at least  $30\text{ W m}^{-2}$  in much of the Sahara, Horn of Africa, South Africa, and Australia, while some regions in the northern high latitudes would decrease approximately  $20\text{ W m}^{-2}$ .

*f. Grid-by-grid comparison of annual averages with reanalysis data*

While our products' large average differences in Fig. 5 with both reanalysis datasets in the Sahara and Middle

East cannot be directly compared with estimates from Jung et al. (2011; who only include vegetated regions), temperature differences may partially account for these differences, as indicated by Fig. 3b in Siemann et al. (2018); this shows over 60% of the uncertainty from the temperature gradient in these regions, as well as the roughness length, as indicated by the drastic sensitivity of sensible heat flux of over  $500\text{ W m}^{-2}$  to changes in roughness length of 1 m seen in Fig. 1 of Siemann et al. (2018). Although the differences in roughness length between the constrained products and the reanalysis in these regions are not likely to be greater than 1 m because it is the bare soil type, even a fraction of 1-m difference could likely cause the estimates to differ by tens of watts per square meter, based on the change of over  $500\text{ W m}^{-2}$  in these regions due to a change in 1 m for roughness length in Fig. 1 of Siemann et al. (2018).

Regarding the positive average biases seen against the reanalysis datasets, Jung et al. (2011) serve as a literature comparison, spatially. As noted in section 4c, the estimates in Jung et al. (2011) are consistent with our constrained product estimates in the northern latitudes of Eurasia and eastern North America. Similar to the positive biases of our constrained products with reanalysis in the northern regions of South America as well as in Indonesia, we note in section 4c that the differences in our products here with Jung et al. (2011) can be attributed to the measurement and validation complications for remotely sensed products in the tropics, as well as the uncertainty in Siemann et al. (2018) due to the uncertainty in  $C_{z0}$ . As for the positive biases in southern Asia, the constrained sensible heat flux estimates of  $30\text{--}60\text{ W m}^{-2}$  in these regions are consistent with the estimates in Jung et al. (2011) of  $30\text{--}50\text{ W m}^{-2}$ . In the case of the central Asia positive biases, the reanalysis values are more consistent with Jung et al.'s (2011) estimates of  $50\text{--}63\text{ W m}^{-2}$  in this region, while our estimates reach  $70\text{--}90\text{ W m}^{-2}$ . This overestimation can be attributed to the uncertainty of  $25\text{--}35\text{ W m}^{-2}$  in this region of Asia seen in Siemann et al. (2018), which is due to both the dominant uncertainty in the temperature gradient in this region seen in Fig. 3 of Siemann et al. (2018) and also the large sensitivities in this region to the roughness length for momentum and the  $C_{z0}$  seen in Fig. 1 of Siemann et al. (2018).

The negative biases seen in central Africa and South America can be similarly attributed to the complications noted above and in section 4c regarding both issues with measurement and validation in the tropics, as well as uncertainties of up to  $15\text{ W m}^{-2}$  in central Africa due to uncertainties in  $C_{z0}$  and the temperature gradient seen in Figs. 2 and 3 of Siemann et al. (2018). As for the negative biases of our constrained products with the



MERRA reanalysis in central South America, our estimates are consistent with Jung et al. (2011): between 30 and  $50 \text{ W m}^{-2}$ . The negative biases of the product using the CFSR temperature gradient with the reanalysis data occur more vastly, spatially, because the products using this air temperature have lower sensible heat fluxes relative to our other constrained (and unconstrained) products due to the differences between the air temperatures, as noted in sections 3c and 3d and as discussed in section 4d.

As for the correlation coefficients, the larger correlation coefficients between 0.6 and 0.8 in regions such as the Sahara, the Arabian Peninsula, and the high northern latitudes could occur despite larger biases because there is not much seasonal variation in these regions, producing smaller changes from day to day, which could occur similarly to those day-to-day changes in the reanalysis, despite different magnitudes of the fluxes. Additionally, the location of the highest correlation coefficients is almost always (with the exception of the Sahara) outside the traditional definition of the tropics of  $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ , where satellite measurements offer more coverage for products such as the HIRS-consistent LST.

If Fig. 5 used products constrained with the net radiation using the SRB or VIC albedos, similar patterns would be seen using the SRB albedo, but the biases in the Sahara and Middle East would be larger using the VIC albedo. Additionally, for products constrained with the net radiation using the SRB and VIC albedos, correlations would be similar, except there would be more negative correlations in the tropics.

#### g. Validation of sensible heat flux against FLUXNET observations

In comparing our products and the reanalysis datasets with observed data in Fig. 6, we also forced our tower algorithm with tower-observed inputs (the Rs\_Tforc\_products) to provide comparison with the best possible performance we can expect from our algorithm, given our model structure. Of the 41 sites that have the required input data for our algorithm, the correlation coefficients for the Rs\_Tforc\_products are expected to exceed all other datasets at these sites because the ground observations should be the truthful meteorological conditions producing Obs\_Tower estimates, as seen in Fig. 6. It should be noted that the number of stations is just under 1/3 of the total number of stations used for the other datasets, so the comparison is not a direct comparison. The positive performance in correlation coefficients at the stations that have enough tower input data demonstrates the improved average and median correlations when using the observed meteorological data, and this positive performance in correlation

coefficients supports our use of this algorithm and constraining approach. This comparison also demonstrates the difference between the results using meteorological forcing datasets that include averages for  $50 \text{ km}^2$  grid cells located over a tower site and the meteorological observations on the ground at that site. The biases of the Rs\_Tforc\_products in Fig. 6b are the highest of the biases, possibly attributed to errors when constraining the tower-forced data with the latent heat flux forced with CFSR meteorological forcings and the three reanalysis-based air temperatures. Despite these inconsistencies, using tower forcings in the unconstrained latent heat flux would be outside the feasible range of this study due to computational and consistency issues within the latent heat flux algorithm inputs. RMSEs show increased performance for the Rs\_Tforc\_products relative to our constrained products and the reanalysis estimates that is consistent with the correlation coefficient performance.

When considering performance among our constrained products, the Rs\_Rnet\_alM\_T\_CFSR and Rs\_Rnet\_alV\_T\_CFSR products have the largest mean and median correlation coefficients, as well as some of the smallest mean and median RMSEs. Both the mean and median average differences for the Rs\_Rnet\_alM\_T\_CFSR product are slightly smaller, compared to the Rs\_Rnet\_alS\_T\_CFSR product, but the Rs\_Rnet\_alV\_T\_CFSR mean and median biases are closer to zero. Despite the smaller mean and median biases of Rs\_Rnet\_alV\_T\_CFSR, the range of the 25th to 75th percentile is smaller for the biases and RMSEs for the Rs\_Rnet\_alM\_T\_CFSR product. Additionally, the tower-forced datasets constrained with the MODIS and SRB albedo have the lower median and mean biases and RMSEs, but the 25th-percentile bias is smaller for the datasets constrained with the MODIS albedo. This information, coupled with the results in the right side of Table 4 that show the products using the MODIS albedo have the closest global averages to the literature values, and coupled with the difference between Rs\_Rnet\_alM\_T\_CFSR and Rs\_Rnet\_alS\_T\_CFSR of  $42.4$  to  $43.9 \text{ W m}^{-2}$ , respectively, demonstrates that the best-performing product is the Rs\_Rnet\_alM\_T\_CFSR, and the best-performing albedo in the net radiation is generally the MODIS albedo. The poor performance at outlier stations due to land-cover heterogeneity is discussed in the discussion of Fig. S3, which shows the outliers.

In Fig. 7, the higher mean and median correlation coefficients in each category of the reanalysis estimates above our constrained products can be explained by any adjustments done to the reanalysis estimates to ensure proper seasonal cycles, which we do not correct for in our constrained products. Although our constrained

products have larger biases at the wetland and water sites, the mean and median RMSEs are lowest and the correlation coefficients are among the highest at these sites. This category only contains two towers, so the averaging for the biases would not drag the mean and median down toward zero as much as it could for categories with more towers, likely having both positive and negative biases among them. Although the shrubland and savanna sites have high mean and median correlation coefficients, they have some of the largest mean and median RMSEs, consistent with the correlations above 0.6 with reanalysis in Australia in Fig. 5, but larger biases with both reanalysis of over  $\pm 5 \text{ W m}^{-2}$  and often up to  $\pm 15 \text{ W m}^{-2}$ . Grassland sites also show large RMSEs and large biases, consistent with the biases of  $5\text{--}15 \text{ W m}^{-2}$  or more with the reanalysis in grassland regions of North America.

When dividing the towers by region, the African and Australian towers with the largest mean and median average biases and the largest mean RMSEs are consistent with the above discussion of the performance of the shrubland and savannas category. Additionally, these categories only have one and five towers, respectively, so the impacts of a few poor-performing towers can manifest stronger in these medians and means. Both our constrained products and reanalysis products perform poorly at the South American towers across all statistics (except for the biases for the reanalysis products), which can be attributed to the issues discussed with the Amazon, as several of the sites are in Brazil, and to the measurements and validation in the tropics in sections 4c and 4f.

Consistency is also seen by climate class for the lowest mean and median correlation coefficients, larger median biases, and large median RMSEs in the tropical towers. The performance of the arid climate class follows that of the shrubland and savannas category discussed above. The poor performance of the polar towers may occur because only three towers are classified as polar towers, making it the smallest category among the climate classes present in this study.

## 5. Conclusions

In this study, we develop 6-hourly,  $0.5^\circ$ , global, terrestrial sensible heat flux products based on remotely sensed data and reanalysis meteorology. Global averages for four of the six unconstrained products overestimate the previously published estimates by at least  $18 \text{ W m}^{-2}$ . Additionally, patchy negative sensible heat flux estimates at high latitudes and in regions of the Congo, as well as large overestimation relative to Jung et al. (2011) in the Midwest CONUS, Great Plains,

Canadian Plains, and central Asia, prompted us to constrain products using the Bowen ratio and available energy at the daily scale. We examine the annual averages, spatial patterns, and seasonal dynamics of these constrained products and compare them with reanalysis data and data from the literature.

The constraining process produces global annual averages that are more consistent with spatial patterns and magnitudes of the estimates in Jung et al. (2011), apart from some regions of the tropics such as the Congo or the Amazon. These remaining inconsistencies in the Congo and the Amazon are attributed to complexities of measurements and lack of validation of remotely sensed products in the tropics, in addition to large uncertainties in the sensible heat flux in the Amazon (Siemann et al. 2018). Negative fluxes in the northern high latitudes are also, appropriately, in the SON and DJF seasonal averages for the constrained products instead of the annual averages. Although all constrained products except Rs\_Rnet\_alM\_G\_CFSR have global averages above the uncertainty range around the  $32 \text{ W m}^{-2}$  average in Wild et al. (2015), they are all within the uncertainty range of  $\pm 7.0 \text{ W m}^{-2}$  from the  $38.8 \text{ W m}^{-2}$  average in L'Ecuyer et al. (2015). Rs\_Rnet\_alM\_T\_CFSR performs closest to the FLUXNET ground observations at the monthly scale, in addition to having one of the closest global annual averages to reported estimates in the literature. We attribute most of the poor-performing outliers in the monthly analysis to the difficulty in representing the true fluxes at sites with heterogeneous terrain within the  $50 \text{ km}^2$  grid cell of our datasets (see the supplemental discussion related to Fig. S3). Of the 154 stations using 6 months minimum of data passing the quality control, our constrained products perform best at the wetland, water, evergreen forests, deciduous forests, and croplands when divided by vegetation type; the North American, European, and Asian sites when divided by region; and the temperate and cold sites when divided by Köppen climate class.

Our sensible heat flux estimates are most useful for benchmarking climate models at the annual or monthly scale. Local or regional validation or benchmarking should be done on a case-by-case basis considering the strengths and weaknesses, including the sensitivities and uncertainties in Siemann et al. (2018), for regions and land-cover types. Improvements in the sensible heat flux can be made with improvements in the quality of the input variables, particularly in the measurement of the temperature gradient, the roughness length dynamics, and the  $C_{z0}$  estimates. Future work could include expanding the consistent temperatures measured from the same sensor, such as NASA's Atmospheric Infrared Sounder (AIRS), into globally consistent products with

the Bayesian methodology used to form the HIRS-consistent LST for use in the temperature gradient driving the sensible heat flux.

**Acknowledgments.** This work is supported by NOAA Grant NA11OAR4310175.

## REFERENCES

- Baldocchi, D., 2008: 'Breathing' of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems. *Aust. J. Bot.*, **56**, 1–26, <https://doi.org/10.1071/BT07151>.
- Bastiaanssen, W. G. M., M. Meneti, R. A. Feddes, and A. A. M. Holtslag, 1998: A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *J. Hydrol.*, **212–213**, 198–212, [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4).
- Beljaars, A. C. M., 1995: The parametrization of surface fluxes in large-scale models under free convection. *Quart. J. Roy. Meteor. Soc.*, **121**, 255–270, <https://doi.org/10.1002/qj.49712152203>.
- Bowen, I. S., 1926: The ratio of heat losses by conduction and by evaporation from any water surface. *Phys. Rev.*, **27**, 779–787, <https://doi.org/10.1103/PhysRev.27.779>.
- Brutsaert, W., 2005: Water aloft: Fluid mechanics of the lower atmosphere. *Hydrology: An Introduction*, Cambridge University Press, 23–78.
- Campbell, G. S., and J. M. Norman, 1998: Radiation basics. *An Introduction to Environmental Biophysics*, Springer, 147–165.
- Chaney, N. W., J. D. Herman, M. B. Ek, and E. F. Wood, 2016: Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *J. Geophys. Res. Atmos.*, **121**, 13 218–13 235, <https://doi.org/10.1002/2016JD024821>.
- Chen, F., J. Janjić, and K. Mitchell, 1997: Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta Model. *Bound.-Layer Meteor.*, **85**, 391–421, <https://doi.org/10.1023/A:1000531001463>.
- Clemen, R. T., and R. L. Winkler, 1999: Combining probability distributions from experts in risk analysis. *Risk Anal.*, **19**, 187–203, <https://doi.org/10.1111/j.1539-6924.1999.tb00399.x>.
- Coccia, G., and E. Todini, 2011: Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrol. Earth Syst. Sci.*, **15**, 3253–3274, <https://doi.org/10.5194/hess-15-3253-2011>.
- , A. L. Siemann, M. Pan, and E. F. Wood, 2015: Creating consistent datasets by combining remotely-sensed data and land surface model estimates through Bayesian uncertainty post-processing: The case of land surface temperature from HIRS. *Remote Sens. Environ.*, **170**, 290–305, <https://doi.org/10.1016/j.rse.2015.09.010>.
- Defries, R. S., M. C. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland, 2000: A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biol.*, **6**, 247–254, <https://doi.org/10.1046/j.1365-2486.2000.00296.x>.
- Dolman, A. J., D. G. Miralles, and R. A. M. de Jeu, 2014: Fifty years since Monteith's 1965 seminal paper: The emergence of global ecohydrology. *Ecohydrology*, **7**, 897–902, <https://doi.org/10.1002/eco.1505>.
- Dye, D. G., 2002: Variability and trends in the annual snow-cover cycle in Northern Hemisphere land areas, 1972–2000. *Hydrol. Processes*, **16**, 3065–3077, <https://doi.org/10.1002/hyp.1089>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction Operational Eta Model. *J. Geophys. Res.*, **108**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Gutman, G., and A. Ignatov, 1998: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int. J. Remote Sens.*, **19**, 1533–1543, <https://doi.org/10.1080/014311698215333>.
- Hansen, M. C., R. S. Defries, J. R. G. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.*, **21**, 1331–1364, <https://doi.org/10.1080/014311600210209>.
- Hengl, T., and Coauthors, 2014: SoilGrids1km—Global soil information based on automated mapping. *PLOS ONE*, **9**, e105992, <https://doi.org/10.1371/journal.pone.0105992>; Corrigendum, **9**, e114788, <https://doi.org/10.1371/journal.pone.0114788>.
- Holtslag, A. A. M., and H. A. R. de Bruin, 1988: Applied modeling of the nighttime surface energy balance over land. *J. Appl. Meteor.*, **27**, 689–704, [https://doi.org/10.1175/1520-0450\(1988\)027<0689:AMOTNS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1988)027<0689:AMOTNS>2.0.CO;2).
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, [https://doi.org/10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).
- Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.*, **116**, D02102, <https://doi.org/10.1029/2010JD014545>.
- Jung, M., and Coauthors, 2011: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological estimates. *J. Geophys. Res.*, **116**, G00J07, <https://doi.org/10.1029/2010JG001566>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kustas, W. P., B. J. Choudhury, M. S. Moran, R. J. Reginato, R. D. Jackson, L. W. Gay, and H. L. Weaver, 1989: Determination of sensible heat flux over sparse canopy using thermal infrared data. *Agric. For. Meteorol.*, **44**, 197–216, [https://doi.org/10.1016/0168-1923\(89\)90017-8](https://doi.org/10.1016/0168-1923(89)90017-8).
- L'Ecuyer, T. S., and Coauthors, 2015: The observed state of the energy budget in the early twenty-first century. *J. Climate*, **28**, 8319–8346, <https://doi.org/10.1175/JCLI-D-14-00556.1>.
- Liang, S., and Q. Liu, 2012: Global land surface products: Albedo product data collection (1985–2010). Beijing Normal University, accessed 6 May 2015, <https://doi.org/10.6050/glass863.3001.db>.
- , K. Wang, X. Zhang, and M. Wild, 2010: Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **3**, 225–240, <https://doi.org/10.1109/JSTARS.2010.2048556>.
- Liu, N. F., Q. Liu, L. Z. Wang, S. L. Liang, J. G. Wen, Y. Qu, and S. H. Liu, 2013: A statistics-based temporal filter algorithm to map spatiotemporally continuous shortwave albedo from MODIS data. *Hydrol. Earth Syst. Sci.*, **17**, 2121–2129, <https://doi.org/10.5194/hess-17-2121-2013>.
- Liu, Q., L. Wang, Y. Qu, N. Liu, S. Liu, H. Tang, and S. Liang, 2013: Preliminary evaluation of the long-term GLASS albedo product. *Int. J. Digital Earth*, **6** (Suppl.), 69–95, <https://doi.org/10.1080/17538947.2013.804601>.

- Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, R. A. M. de Jeu, W. Wagner, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk, 2012: Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.*, **123**, 280–297, <https://doi.org/10.1016/j.rse.2012.03.014>.
- Lu, S., T. Ren, and R. Horton, 2007: An improved model for predicting soil thermal conductivity from water content at room temperature. *Soil Sci. Soc. Amer. J.*, **71**, 8–14, <https://doi.org/10.2136/sssaj2006.0041>.
- Mecikalski, J. R., G. R. Diak, M. C. Anderson, and J. M. Norman, 1999: Estimating fluxes on continental scales using remotely sensed data in an atmospheric–land exchange model. *J. Appl. Meteor.*, **38**, 1352–1369, [https://doi.org/10.1175/1520-0450\(1999\)038<1352:EFOCSU>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1352:EFOCSU>2.0.CO;2).
- Miralles, D. G., and Coauthors, 2016: The WACMOS-ET project—Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrol. Earth Syst. Sci.*, **20**, 823–842, <https://doi.org/10.5194/hess-20-823-2016>.
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the atmosphere near the ground. *Tr. Geofiz. Inst., Akad. Nauk SSSR*, **24**, 163–187.
- Monteith, J. L., 1973: *Principles of Environmental Physics*. Edward Arnold, 241 pp.
- Murray, T., and A. Verhoef, 2007: Moving towards a more mechanistic approach in the determination of soil heat flux from remote measurements: I. A universal approach to calculate thermal inertia. *Agric. For. Meteorol.*, **147**, 80–87, <https://doi.org/10.1016/j.agrformet.2007.07.004>.
- Norman, J. M., W. P. Kustas, J. H. Prueger, and G. R. Diak, 2000: Surface flux estimation using radiometric temperature: A dual-temperature-difference method to minimize measurement errors. *Water Resour. Res.*, **36**, 2263–2274, <https://doi.org/10.1029/2000WR900033>.
- Oak Ridge National Laboratory, 2016: Fluxnet datasets list. Oak Ridge National Laboratory Distributed Active Archive Center, accessed 8 August 2013, <http://fluxnet.ornl.gov>.
- Ogunjemiyo, S. O., S. K. Kaharabata, P. H. Schuepp, I. J. MacPherson, R. L. Desjardins, and D. A. Roberts, 2003: Methods of estimating CO<sub>2</sub>, latent heat and sensible heat fluxes from estimates of land cover fractions in the flux footprint. *Agric. For. Meteorol.*, **117**, 125–144, [https://doi.org/10.1016/S0168-1923\(03\)00061-3](https://doi.org/10.1016/S0168-1923(03)00061-3).
- Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432, <https://doi.org/10.2151/jmsj.85.369>.
- Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.*, **9**, 857–861, [https://doi.org/10.1175/1520-0450\(1970\)009<0857:TMROWS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2).
- Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644, <https://doi.org/10.5194/hess-11-1633-2007>.
- Pipunic, R. C., J. P. Walker, and A. Western, 2008: Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sens. Environ.*, **112**, 1295–1305, <https://doi.org/10.1016/j.rse.2007.02.038>.
- Qu, Y., Q. Liu, S. Liang, L. Wang, N. Liu, and S. Liu, 2014: Direct-estimation algorithm for mapping daily land-surface broadband albedo from MODIS data. *IEEE Trans. Geosci. Remote Sens.*, **52**, 907–919, <https://doi.org/10.1109/TGRS.2013.2245670>.
- Ren, H., S. Liang, G. Yan, and J. Cheng, 2013: Empirical algorithms to map global broadband emissivities over vegetated surfaces. *IEEE Trans. Geosci. Remote Sens.*, **51**, 2619–2631, <https://doi.org/10.1109/TGRS.2012.2216887>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Sellers, P. J., and Coauthors, 1997: BOREAS in 1997: Experiment overview, scientific results, and future directions. *J. Geophys. Res.*, **102**, 28 731–28 769, <https://doi.org/10.1029/97JD03300>.
- Sellers, D., 1965: *Physical Climatology*. The University of Chicago Press, 272 pp.
- Serreze, M. C., and Coauthors, 2000: Observational evidence of recent change in the northern high-latitude environment. *Climatic Change*, **46**, 159–207, <https://doi.org/10.1023/A:1005504031923>.
- Sheffield, J., E. F. Wood, and F. Munoz-Arriola, 2010: Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data. *J. Hydrometeorol.*, **11**, 253–275, <https://doi.org/10.1175/2009JHM1176.1>.
- Siemann, A. L., G. Coccia, M. Pan, and E. F. Wood, 2016: Development and analysis of a long-term, global, terrestrial land surface temperature dataset based on HIRS satellite retrievals. *J. Climate*, **29**, 3589–3606, <https://doi.org/10.1175/JCLI-D-15-0378.1>.
- , N. Chaney, and E. F. Wood, 2018: Sensitivity and uncertainty of a long-term, high-resolution, global, terrestrial sensible heat flux dataset. *J. Geophys. Res. Atmos.*, <https://doi.org/10.1029/2017JD027785>, in press.
- Smith, A., N. Lott, and R. Vose, 2011: The Integrated Surface Database: Recent developments and partnerships. *Bull. Amer. Meteor. Soc.*, **92**, 704–708, <https://doi.org/10.1175/2011BAMS3015.1>.
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, T. Zhang, J. C. Mikovitz, and L. M. Hinkelman, 2011: 24.5-year data set released. *GEWEX News*, Vol. 21, No. 1, International GEWEX Project Office, Silver Spring, MD, 10–12, [https://www.gewex.org/gewex-content/files\\_mf/1432209318Feb2011.pdf](https://www.gewex.org/gewex-content/files_mf/1432209318Feb2011.pdf).
- Stephens, G. L., and Coauthors, 2012: An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.*, **5**, 691–696, <https://doi.org/10.1038/ngeo1580>.
- Stone, M., 1961: The opinion pool. *Ann. Math. Stat.*, **32**, 1339–1342, <https://doi.org/10.1214/aoms/1177704873>.
- Timmermans, W. J., W. P. Kustas, M. C. Anderson, and A. N. French, 2007: An intercomparison of the surface energy balance algorithm for land (SEBAL) and the two-source energy balance (TSEB) modeling schemes. *Remote Sens. Environ.*, **108**, 369–384, <https://doi.org/10.1016/j.rse.2006.11.028>.
- Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manage.*, **6**, 123–137, <https://doi.org/10.1080/15715124.2008.9635342>.
- Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, **90**, 311–323, <https://doi.org/10.1175/2008BAMS2634.1>.

- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, <https://doi.org/10.1256/qj.04.176>.
- Vinukollu, R. K., E. F. Wood, C. R. Ferguson, and J. B. Fisher, 2011: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sens. Environ.*, **115**, 801–823, <https://doi.org/10.1016/j.rse.2010.11.006>.
- Wagner, W., W. Dorigo, R. de Jeu, D. Fernandez, J. Benveniste, E. Haas, and M. Ertl, 2012: Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *Proc. 22nd Int. Society for Photogrammetry and Remote Sensing Congress*, Melbourne, Australia, International Society for Photogrammetry and Remote Sensing, 315–321.
- Wang, A., and X. Zeng, 2013: Development of global hourly 0.5° land surface air temperature datasets. *J. Climate*, **26**, 7676–7691, <https://doi.org/10.1175/JCLI-D-12-00682.1>.
- Wang, K., and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.*, **50**, RG2005, <https://doi.org/10.1029/2011RG000373>.
- Wild, M., and Coauthors, 2015: The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dyn.*, **44**, 3393–3429, <https://doi.org/10.1007/s00382-014-2430-z>.
- Ye, H., H.-R. Cho, and P. E. Gustafson, 1998: The changes in Russian winter snow accumulation during 1936–83 and its spatial patterns. *J. Climate*, **11**, 856–863, [https://doi.org/10.1175/1520-0442\(1998\)011<0856:TCIRWS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<0856:TCIRWS>2.0.CO;2).
- Zilitinkevich, S. S., 1995: Non-local turbulent transport: Pollution dispersion aspects of coherent structure of convective flows. *Air Pollution Theory and Simulation*, H. Power, N. Moussiopoulos, and C. A. Brebbia, Eds., Vol. 1, *Air Pollution III*, Computational Mechanics Publications, 53–60.