Physically Plausible Methods for Projecting Changes in Great Lakes Water Levels under Climate Change Scenarios

BRENT M. LOFGREN AND JONATHAN ROUHANA[a]

NOAA/Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan

ABSTRACT

A method for projecting the water levels of the Laurentian Great Lakes under scenarios of human-caused climate change, used almost to the exclusion of other methods in the past, relies very heavily on the large basin runoff model (LBRM) as a component for determining the water budget for the lake system. This model uses near-surface air temperature as a primary predictor of evapotranspiration (ET); as in previous published work, it is shown here that the model's very high sensitivity to temperature causes it to overestimate ET in a way that is greatly at variance with the fundamental principle of conservation of energy at the land surface. The traditional formulation is characterized here as being equivalent to having several suns in the virtual sky created by LBRM. More physically based methods show, relative to the traditional method, often astoundingly less potential ET and less ET, more runoff from the land and net basin supply for the lake basins, and higher lake water levels in the future. Using various methods of estimating the statistical significance, it is found that, at minimum, these discrepancies in results are significant at the 99.998% level. The lesson for the larger climate impact community is to use caution about whether an impact is forced directly by air temperature itself or is significantly forced by season or latitude independently of temperature. The results here apply only to the water levels of the Great Lakes and the hydrology of its basin and do not affect larger questions of climate change.

1. Introduction

The Laurentian Great Lakes of North America contain approximately 20% of the world's surface freshwater. They support intrinsic and commercial interest for shoreline property owners, ecosystems in the lakes themselves and coastal wetlands, commercial and recreational fishing, boating, shipping, and tourism. They are a binational resource of the United States and Canada, and over 30 million people live within their drainage basin (Government of Canada and U.S. EPA 1995).

The issue of the potential influence of future climate change on the water levels of the Great Lakes has been a topic of a significant body of literature. Many examples have used the method pioneered by Croley (1990). This method is based on the use of a suite of regional hydrological models with historically observed input variables altered by ''change factors'' applied to these variables based on general circulation model (GCM) results—either adding the difference between future and present GCM output or multiplying by the ratio between future and present output, depending on the variable type. A partial list of this class of studies includes Hartmann (1990), Chao (1999), Lofgren et al. (2002), and Angel and Kunkel (2010). The set of results using this method has been featured prominently in media coverage and in the assessment and synthesis literature (e.g., Great Lakes Regional Assessment Group 2000; Kling et al. 2003; International Lake Ontario–St. Lawrence River Study Board 2006; Hayhoe et al. 2010).

Other studies have used different approaches to examine the hydrologic budget of the Great Lakes basin. Some have used direct analysis of the GCM-based regional water budget (precipitation minus evaporation) in the region (Manabe et al. 2004; Milly et al. 2005), while others have examined the convergence of water vapor flux in the atmosphere as indicating net water

[a] Current affiliation: University of Notre Dame, South Bend, Indiana.

Corresponding author address: Brent M. Lofgren, NOAA/Great Lakes Environmental Research Laboratory, 4840 S. State Rd., Ann Arbor, MI 48108-9719.
E-mail: brent.lofgren@noaa.gov

budget (Kutzbach et al. 2005). Still others have used regional climate modeling to downscale GCM results and project climate both in terms of the atmosphere and the surface (MacKay and Seglenieks 2013; Lofgren 2014; Notaro et al. 2015). Most of these studies did not include lake level explicitly as an output, although MacKay and Seglenieks (2013) and Notaro et al. (2015) do have such projections based on a limited range of GCM inputs. The group of studies that has not been based on the methodology of Croley (1990) has generally not shown large trends toward reduced water budgets in the Great Lakes, and even though not always explicitly shown, this necessarily implies small changes in lake levels.

Several studies have called into question the use of temperature-based methods for projecting evapotranspiration (ET) in climate change scenarios (e.g., Hobbins et al. 2008; Shaw and Riha 2011; McAfee 2013). Specifically for the Great Lakes region, Lofgren et al. (2011, hereafter LHW) attempted to expose some serious problems with the methodology of Croley (1990) for projection of Great Lakes water budgets and levels under climate change scenarios. These specifically dealt with the large basin runoff model (LBRM), which is the component of the modeling system that simulates the land portion of the Great Lakes drainage basin. The central argument was that the LBRM is overly reliant on near-surface air temperature as a predictor of ET, to the exclusion of the surface energy budget as a predictor; although the LBRM is calibrated such that potential evapotranspiration is constrained by incoming solar radiation during the calibration period, this constraint is not used when considering periods with altered climate regimes. This was broken down into three intersecting subsidiary lines of argument:

1) The ET projected for future climate scenarios using input from GCMs was much greater than that directly simulated by the driving GCMs. The equivalent energy of the difference, in terms of latent heat flux, was shown to be around $30\,\mathrm{W\,m^{-2}}$, a significant discrepancy in the surface energy budget.
2) During periods of observation at land-based flux measurement stations, the near-surface air temperature and ET were decomposed into different temporal classes—the annual cycle, variations with a time scale greater than a month but not annual, and variations with a time scale less than a month. For the annual time scale, there was a very high correlation between air temperature and ET, while for the other time scales, there was almost no correlation. This indicates that the seasonal cycle (i.e., solar energy input) is the major driver of both air temperature and ET at the annual scale, while air temperature

is a poor predictor of ET variability on other time scales. Therefore, thinking of air temperature as either an actual driver of ET or as a proxy that is universally applicable is highly problematic.
3) The potential ET (PET), proportional to what is called by Croley (1983) the ''energy available for evapotranspiration,'' is greater by very large factors in the future scenarios than in the historical base case. For example, a 10-fold increase was shown for all of the land in the Lake Superior basin, which can be taken as equivalent to an increase in the incoming sunlight by that factor. These very large increases were due to LBRM being calibrated such that PET increased in many locations by 30%–50% $(^\circ\mathrm{C})^{-1}$, and even more in a few locations. This contrasts with the point of comparison of the Clausius–Clapeyron (CC) relation, which has an increase in saturated water vapor pressure of 7% $(^\circ\mathrm{C})^{-1}$, and literature sources such as Held and Soden (2006) and Lorenz and DeWeaver (2007), which show increases in large-scale ET under climate change less than 7% $(^\circ\mathrm{C})^{-1}$.

These arguments say that the methods that have been repeatedly used over more than two decades strongly overestimate future ET. In addition to this, LHW gave a small-scale demonstration of what effect these problems might have on projected lake levels. As a simple alternative that fit easily into the framework of LBRM, they devised the energy adjustment (EA) method, in which the changes in PET from a future climate scenario relative to the historical period were given by the ratio of net surface radiation, rather than the exponential function of the air temperature change, dubbed the temperature adjustment (TA) method. Using two different GCM runs as input, this resulted in discrepancies in the level of Lake Michigan–Huron of around 1 m.

These results received very guarded recognition in the 2014 National Climate Assessment (Pryor et al. 2014; Walsh et al. 2014). Some National Climate Assessment authors suggested that what was missing was an assessment of the statistical significance of the difference between the traditional TA method and the more physically based EA method by using a wider selection of GCM data as input (D. Wuebbles and K. Hayhoe 2013, personal communication). Therefore, that is a major goal of the present study. We also introduce two more alternative formulations for PET under climate change in addition to the already-used EA method, as additional points of comparison to the TA method. Arguments 1 and 2 from LHW, in the list above, stand as already stated there, but our further use of many more GCMs as input leads to further underline argument 3.

The rest of this paper is organized as follows. Section 2 summarizes the methodologies that are used here. Section 3 presents the results, including formal analysis of the statistical significance of the discrepancy among these alternative methods. Section 4 discusses the results and the way that strikingly extreme intermediate results within a model can be disguised in the final results. Conclusions are presented in section 5.

2. Methods and experimental design

a. The LBRM

The LBRM is the model used for simulation of the land portion of the Great Lakes basin. Key features of the formulation of LBRM for purposes of this study include the determination of "energy available for evapotranspiration" (for convenience, often referred to here as PET; see further discussion below) and the proportion of PET that is manifested as ET. As stated in Croley (1983), because of considerations of data availability, an a priori constraint in developing LBRM was that only daily air temperatures and precipitation be used as input data.

A separate set of nine parameters for the LBRM are calibrated for each subbasin of the Great Lakes basin, in order to minimize the objective function of root-mean-square error between the simulated river outflow from that subbasin and the measured outflow. Outflow data are from U.S. Geological Survey flow gauges and are scaled from the portion of the subbasin that they measure to the entire subbasin by the area ratio method (Croley and Hartmann 1984). No other observed quantities (e.g., soil moisture, pan evaporation, water table depth, and inland lake or wetland levels) enter into the objective function used for calibration. The parameter search algorithm was a simple recursive downslope search. That is, if a parameter is adjusted to a new test value and it improves the objective function, then adjust it more in the same direction and test again; otherwise, adjust in the opposite direction and eventually switch to adjusting another parameter. Continue cycling through the full set of parameters until all converge to constant values. The parameters have been recalibrated since Croley (1983); however, this is not fully documented, but the full set of calibrated parameters for each subbasin in LBRM, under the current calibration, is listed in Table 1 of Lofgren and Rouhana (2016). A brief description of the function of those parameters is in the caption of that table, with more details in Croley (1983).

The key calibrated parameter for the purposes of this study is the base temperature T_b. Table 1 of Lofgren and Rouhana (2016) shows that this has values among the subbasins ranging from 1.1° to 14.0°C, with most values between

about 2.3° and 8°C; the lower values occur most often in the subbasins of Lake Superior. PET is assumed to take the form Ae^{T/T_b}. During the calibration process, the annual total of PET is taken to be such that its latent heat equivalent is equal to the annual incoming solar radiative energy:

$$LA \sum^{\text{year}} e^{T/T_b} = \sum^{\text{year}} S, \tag{1}$$

where L is the latent heat of vaporization per unit mass, A is a parameter (cm day^{-1} or other units of evapotranspiration), T is daily mean air temperature, and S is the top-of-atmosphere incident solar radiation. The summation in Eq. (1) uses a daily time step for the snow-free portion of the year as, when snow is present, ET is assumed to be zero, and a separate formulation is used for snowmelt (Croley 1983). In the calibration process, for each test value of T_b, the value of A is uniquely determined using Eq. (1). Thus, Eq. (1) constrains the total PET over the year, while T_b controls its distribution throughout the year and its fluctuation with other modes of temperature variability. It should be noted, however, that outside of the calibration process, including in future projections using adjustments to the temperature based on GCM results, the constraint of Eq. (1) is not enforced. That is, the PET increases by a factor of $e^{\Delta T/T_b}$ corresponding to the change in temperature between current and future conditions, equivalent to the solar radiation also increasing by the same factor.

Actual ET is based on a complementary ET scheme similar to that of Morton (1983). ET has two components, from the upper soil zone and lower soil zone, with the upper soil zone's ET formulated as

$$ET_U = Ae^{T/T_b} \frac{\alpha_U USZM}{1 + \alpha_U USZM + \alpha_L LSZM}, \tag{2}$$

and similarly for the lower soil zone,

$$ET_L = Ae^{T/T_b} \frac{\alpha_L LSZM}{1 + \alpha_U USZM + \alpha_L LSZM}, \tag{3}$$

where USZM and LSZM are soil moisture content in the upper and lower soil zone, respectively, and α_U and α_L are calibrated parameters, referred to in Table 1 of Lofgren and Rouhana (2016) as parameters 4 and 7, respectively. It should be noted that the calibrated values of α_U range between 1.1×10^{-10} and 9.1×10^3, while α_L ranges from 1.0×10^{-10} to 3.2×10^3 (all in units of cm^{-1}), with neither a rigorous nor intuitive physical rationale for these extremely wide ranges of value, but simply that these values result from the calibration.

To clarify the difference in nomenclature between this paper and Croley (1983), we here use the term

"potential evapotranspiration" to refer to the quantity Ae^{T/T_b}, because it is the variable in LBRM that best corresponds to the intuitive and widely used definition of PET as the amount of ET that would occur when unlimited soil moisture is available. The sum of Eqs. (2) and (3) is the total ET and asymptotically approaches Ae^{T/T_b} as lower soil zone moisture tends toward infinity (USZM has an uncalibrated a priori upper limit of 2 cm, with any excess over this being routed as surface runoff).

When multiplied by L, Ae^{T/T_b} is referred to in Croley (1983) as "energy available for evapotranspiration." Croley uses the term PET to refer to the water equivalent of the portion of the energy available for ET that does not actually result in ET, that is, the quantity when you substitute 1 into the numerator of the fraction in either Eq. (2) or Eq. (3). Thus, the sum of the two components of ET plus the PET (under Croley's definition) are proportional to the energy available for ET:

$$\mathrm{ET}_U + \mathrm{ET}_L + \mathrm{PET}_{\mathrm{Croley}}$$
$$= Ae^{T/T_b}\left(\frac{\alpha_U \mathrm{USZM}}{1 + \alpha_U \mathrm{USZM} + \alpha_L \mathrm{LSZM}} + \frac{\alpha_L \mathrm{LSZM}}{1 + \alpha_U \mathrm{USZM} + \alpha_L \mathrm{LSZM}} + \frac{1}{1 + \alpha_U \mathrm{USZM} + \alpha_L \mathrm{LSZM}}\right) = Ae^{T/T_b}. \quad (4)$$

b. Experimental design

The latent heat of ET is one term in the surface energy budget of the earth, and fundamental physical laws require a balance among all of the terms of this budget. This energy budget constraint lies explicitly behind diagnostic formulations of PET, such as that of Priestley and Taylor (1972). It is also the basis for surface components of climate models, with an early example being that of Manabe and Holloway (1975), with evolution to more complex versions, such as Dickinson et al. (1986) and its newer descendants. In these schemes, net surface radiative energy, as a major term in the surface energy budget, is a primary driver of PET; it is not built into the basic formulation of the LBRM outside of its calibration period.

Even most sources of information directed at the general public refrain from directly saying that the increased capacity for water vapor in warmer air causes increased evapotranspiration, but they do often mention these concepts in close proximity to each other (e.g., Brean 2015). Therefore, there can tend to be a simplifying assumption that potential ET will increase with temperature in proportion to the CC relationship. An analysis related to this concept is found in d'Orgeville et al. (2014), in which they compare regional precipitation changes associated with climate change simulations to the "expected" increases, assuming proportionality to the CC relation. They did, in fact, find that in the Great Lakes basin, their regional model showed an increase in precipitation that agreed well with this assumption, approximating 7% (°C)$^{-1}$.

The general framework of previous climate change sensitivity experiments using LBRM and the other associated models under the method of Croley (1990) was to have one run, the base case, which simply used station-based meteorological observations as drivers,

for the period 1948–2005 in our case. Then, to represent future time periods, the same historical meteorology was used as the foundation of the inputs to LBRM, but was adjusted to values intended to represent the changed climate, based on results from GCMs. Croley (1990) notes that this "change factor" method was used because of instructions from the U.S. Environmental Protection Agency, the funding agency for that study. Along with this traditional method of treating climate change in the LBRM (here called the TA method), based on its embedded assumptions about the drivers of ET, we will be using three additional methods of applying GCM output to the LBRM for evaluation of response to climate scenarios (summarized in Table 1).

1) The EA method, as in LHW, uses the sum of the sensible heat flux and latent heat flux given by the GCM as an indication of the net radiative energy flux at the surface. The ratio of the GCM's value in the future time period to the historical time period is used as a multiplier for the LBRM's PET from the base case.

2) The EA method of LHW got some legitimate criticism from Scheff and Frierson (2014) for ignoring the term that is directly driven by air temperature in formulations such as that of Priestley and Taylor (1972). The Priestley–Taylor (PT) formulation is similar to that of Monteith (1973), also known as Penman–Monteith, but uses an empirical constant multiplier α in place of an additional term for water vapor deficit. We adopt the concept of PT here, where, in addition to multiplying PET by the ratio of surface radiative energy as in the EA method, we also multiply it by a factor directly driven by the air temperature change (to be derived below). This latter factor, however, is between about 1.5% and 4% (°C)$^{-1}$, far less sensitive to temperature than the LBRM under the TA method.

TABLE 1. Description of alternative methods of incorporating GCM-based climate projections into the LBRM. Note that the temperature sensitivity range of PET given for the TA method is the mathematical inverse of the range of values of the temperature parameter T_b.

Name	Abbreviation	Description	Temperature sensitivity of PET
Temperature adjustment	TA	Air temperatures adjusted by GCM climate projections are inserted directly into LBRM's equations for PET	7%–90% $(°C)^{-1}$ depending on subbasin, most between 15% and 50%
Energy adjustment	EA	PET values from LBRM in the base case climate are multiplied by the ratio of net radiation at the surface in the future case vs the base case	No explicit dependence on temperature, only energy budget, but ranges among GCM cases from 0.4% to 4.8% $(°C)^{-1}$
Priestley–Taylor	PT	The energy adjustment method is applied first, then PET is further multiplied by the temperature-dependent factor from the Priestley–Taylor equation [Eq. (5)]	Explicit temperature dependence ranging between 1.2% $(°C)^{-1}$ (at 30°C) and 4.0% $(°C)^{-1}$ (at 0°C), plus dependence on energy budget
Clausius–Clapeyron	CC	PET from the base case is multiplied by a factor representing the change in water vapor capacity associated with the temperature change from the GCM	7% $(°C)^{-1}$

3) The causal relationship between ET and atmospheric water holding capacity is not direct and has been a source of confusion, but the CC relation seems to set a maximum bound on sensitivity of ET to climate change. So we use it as an additional point of comparison, multiplying the PET from the base case by the change in water vapor holding capacity of the atmosphere (expressed as a ratio) to project future PET. Even though this method also appears to overestimate the sensitivity of PET to climate change (based on, e.g., Held and Soden 2006; Lorenz and DeWeaver 2007), we regard it as "less wrong" than the TA method and therefore find it useful as another point of comparison to demonstrate the problems that exist in the TA method.

We approximate the CC relationship in an exponential form, giving an approximate saturation vapor pressure of

$$e_s = (0.632 \text{kPa}) e^{T/15.51°C}. \tag{5}$$

This formula is fitted so that it gives the exact values from the table in Lide (2005) at 5° and 25°C. In the range of 0°–30°C, this simplified formula lies within 3.4% of the actual value. Under the PT formulation, further calculations detailed in Lofgren and Rouhana (2016, this portion corrected in the revised version) yield the following sensitivity directly attributable to air temperature:

$$\frac{\partial \text{PET}/\partial T}{\text{PET}} = \frac{4.32 \times 10^{-3}}{0.0407 e^{T/15.51°C} + 0.067}. \tag{6}$$

This expression has units of inverse degrees Celsius, so when multiplied by the change in air temperature, it yields a dimensionless multiplier to be applied to PET.

c. Other model components

In addition to the runoff from the land in the lakes' basins, the other drivers of lake level are the overlake precipitation and evaporation. As in Angel and Kunkel (2010) and preceding studies, overlake precipitation was taken from Thiessen polygon weighting of data from land-based stations for the base case (on the order of 100 stations and polygons per lake, with varying numbers available over time) and adjusted by ratios of GCM precipitation for the future cases. Also, as in previous studies, evaporation from the lakes was calculated using the large lake thermodynamics model (LLTM; Croley 1989), which projects water temperatures of the Great Lakes by treating them as lumped one-dimensional lakes. It estimates solar and longwave radiation exchanges at the surface based on cloud cover, and evaporation (latent heat flux) and sensible heat flux based on bulk aerodynamic formulas. Lake temperature profiles are then calculated using this energy budget and formulas for vertical mixing.

Since the overlake precipitation and evaporation are not affected by the method of adjusting the LBRM for climate change, they have only one value for each GCM realization. They are combined with the LBRM results for each of the methods (TA, EA, PT, and CC) to get net basin supply (NBS) values that correspond to each GCM realization and each method—runoff from LBRM, plus overlake precipitation, minus overlake evaporation from LLTM.

These net basin supplies are then the input to the Coordinated Great Lakes Routing and Regulation Model (CGLRRM), developed by the U.S. Army Corps of Engineers and Environment Canada (now Environment

TABLE 2. Summary of GCM runs used. In the columns under the RCPs, the number of model realizations using that RCP is indicated. All of these cases were used with the lowest-numbered model run from the historical time period.

Institution and country	Model name and reference	RCP 4.5	RCP 6.0	RCP 8.5
National Center for Atmospheric Research, United States	CESM1(CAM5) (Hurrell et al. 2013)	3	3	3
Centre National de Recherches Météorologiques, France	CNRM-CM5 (Voldoire et al. 2013)	1	0	1
Commonwealth Scientific and Industrial Research Organisation, Australia	CSIRO Mk3.6.0 (Rotstayn et al. 2012)	2	2	2
Hadley Centre for Climate Science and Services, United Kingdom	HadGEM2-AO (Collins et al. 2011; Martin et al. 2011)	1	1	0
L'Institut Pierre-Simon Laplace, France	IPSL-CM5A-LR (Dufresne et al. 2013)	2	1	1
Geophysical Fluid Dynamics Laboratory, United States	GFDL CM3 (Delworth et al. 2006; Donner et al. 2011)	1	1	1
Geophysical Fluid Dynamics Laboratory, United States	GFDL-ESM2G (Dunne et al. 2012, 2013)	1	1	1
Geophysical Fluid Dynamics Laboratory, United States	GFDL-ESM2M (Dunne et al. 2012, 2013)	1	1	1

and Climate Change Canada; Tolson 2009). Each set of net basin supply time series was input to the CGLRRM to derive lake levels for Lake Superior, the combined Lake Michigan–Huron, Lake St. Clair (not shown), and Lake Erie (not shown). For reasons related to frequent departures of actual regulation practice from official regulation plans, CGLRRM is not regarded as reliable for simulating Lake Ontario levels, and we have no alternative method.

The values for all variables generated by the GCMs are spatially interpolated to the locations relevant for the overlake precipitation estimate, and for simulation by LBRM and LLTM, by simple inverse distance weighting from the nearest four GCM grid points, without consideration for such factors as elevation or location relative to observing stations with special weather characteristics (e.g., lake breeze or lake-effect precipitation). We understand that this is not a state-of-the-art way of downscaling GCM output, but we want to replicate the methodology of Angel and Kunkel (2010) and predecessor studies in order to isolate the effects of the portions of the formulation that we expect to be even more problematic than the details of downscaling.

d. GCMs

A summary of the GCM runs that were used is in Table 2. This group of model realizations was chosen based on the availability of the data needed for simulations using our modeling system. Our criterion for choosing the first five models listed was that we used the models from phase 5 of the Coupled Model Intercomparison Project (CMIP5), for which the following variables were available on a monthly basis from the Earth System Grid Federation server at Lawrence Livermore National Laboratory for a historical period including

1986–2005 in a single file for each variable and for future periods including 2056–2100 in a single file for each variable: air temperature at 2 m height, humidity at 2 m height, precipitation, cloud cover, sensible heat flux, latent heat flux, and wind at 10 m height (either as a scalar quantity or vector components). We also accessed the Earth System Grid Federation server at the Geophysical Fluid Dynamics Laboratory (GFDL) to access the models generated by GFDL; these output files were in 5-yr increments, requiring additional data handling. It would have been possible to manipulate additional GCM runs, but this set was sufficient for the goals of this study. No other criteria were used to select GCMs, such as their accuracy in historically simulating the climate of the region. The monthly data from these GCMs were spatially interpolated to the lakes and subbasins as in previous studies using the method of Croley (1990).

For each of the GCM realizations, we calculated the climatological values of all variables for each month of the year from the period 2056–75 (referred to as the mid-twenty-first century) and 2081–2100 (referred to as the late twenty-first century) for comparison to the historical period 1986–2005. For the historical period, we used only the lowest-numbered model realization that was available on the data server. So with the 32 model realizations shown in Table 2 and the two time periods, we had 64 scenarios to use as input to the hydrologic model suite.

3. Results

All box plots shown here use the default algorithm for creating box plots using the R software package. Each "data point" represents the average of the variables over the simulation period and over the lake's basin
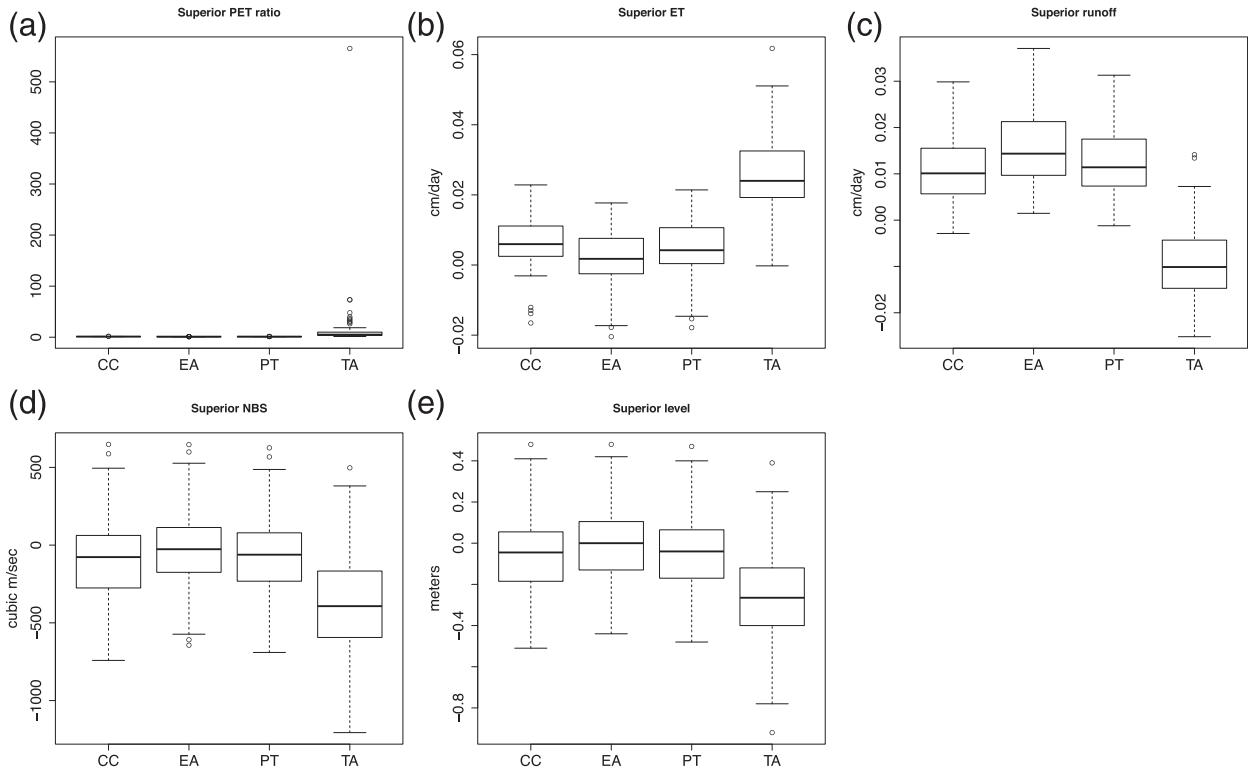
FIG. 1. Box-and-whisker plots of variables averaged over the Lake Superior basin and over the entire year, generated by GCM outputs in combination with LBRM using the TA, CC, PT, and EA methods. (a) Ratio of PET for future and historical time periods (land only), (b) ET for future minus historical time periods (cm day^{-1}; land only), (c) runoff for future minus historical time periods (cm day^{-1}), (d) net basin supply for future minus historical time periods (m^3 s^{-1}), and (e) lake level for future minus historical time periods (m). Each data point used in creating the plots represents an individual GCM realization.

(only the land portion of the basin for PET, ET, and runoff). The boxes extend from the 25% quantile to the 75% quantile level, with a horizontal line indicating the median value. The whiskers extend to the data points that are the farthest away from the box, both above and below, but still within 1.5 times the height of the box. Additional data points outside of this maximum length of the whiskers (outliers) are indicated by individual circles.

a. Lake Superior basin

Lake Superior is the lake whose outflow eventually proceeds to all of the other Great Lakes. Hence, its basin's water budget can significantly influence the lake levels of all of the lakes. As a result of the vagaries of the calibration process of LBRM, the calibrated values of T_b for the subbasins of Lake Superior are generally smaller than those for the other lakes, with many falling in the range of 2°–5°C (contrasting with the Clausius–Clapeyron equivalent value of 15.51°C), and the most extreme low value being 1.1°C in the basin of the Montreal River at the eastern end of Lake Superior.

The first results that we show are the most extreme—the results for changes in PET in the land portion of the Lake Superior basin. These are shown as a ratio of the value in the projected time period to the base case in Fig. 1a. Because the figure's scale is so large to accommodate the changes that result from the TA method, the results for the CC, PM, and EA cases are almost indistinguishable; they cluster slightly above the value of unity, with just a few values below unity. The largest value among these three methods is 1.72 in the case of the CC method for the RCP 8.5 run of the GFDL CM3 for the late twenty-first century.

In sharp contrast to the CC, PM, and EA methods' values for PET are the values under the TA method in the Lake Superior basin (land only). The median value is 5.05, with several values far more extreme than this. The highest value that is shown in the figure is 565.4, again for the RCP 8.5 run of the GFDL CM3 for the late twenty-first century. Although the median increase by a factor of 5.05 may seem small compared to the extreme value of 565.4, it is in fact shockingly large when one considers that it implies the equivalent of an increase in

the incoming solar radiation by a factor of 5.05 [since Eq. (1) links PET in the calibration period to incoming solar radiation].

The actual ET is constrained not only by the PET, but also by the amount of moisture that is available—precipitation and its storage as soil moisture. Therefore, the results for ET among the different methods scale more reasonably to be displayed on the same graph, but still differ markedly. The differences in ET between the base case and climate change cases under the various methods are shown in Fig. 1b (for land only). The median change in ET in the Lake Superior basin under the TA method is $0.024\,\text{cm day}^{-1}$, a 20.9% increase over the base case value of $0.115\,\text{cm day}^{-1}$. Under the PT method, ET is increased by only $0.004\,\text{cm day}^{-1}$. The fact that this increase is much less than the factor of increase in PET is reflective of the fact that both potential and actual ET from land are small during the winter, when ET is energy limited, while the summer season's ET tends to be more moisture limited, especially when PET is increased. Thus, the proportional increase in actual ET is much less than in PET. Nevertheless, the median value of ET increase under the TA method is greater than the top extent of the whiskers for all three of the other methods. Also, the CC, PM, and EA methods' results are much more similar to each other than to the TA method. The difference of $0.020\,\text{cm day}^{-1}$ between TA and PT is the equivalent of $5.8\,\text{W m}^{-2}$—not a really large difference in the surface energy budget, but we need to remember that this is averaged over the entire year and is limited by the availability of moisture, and that the total present direct forcing by CO_2 is around $1.9\,\text{W m}^{-2}$ (http://www.esrl.noaa.gov/gmd/aggi/aggi.html).

In terms of runoff (Fig. 1c), the results are reflective of those for actual ET, but also reflect changes in precipitation. Because there are some GCM runs that project large increases in precipitation, but few project significant decreases, the whiskers extending downward are relatively short, especially for the CC, PT, and EA cases. The entire box for the TA method is below the whiskers for the other three cases, although the whisker at the high end for the TA method extends into the range of the other boxes. This means that, even under the TA method, the GCM cases with larger increases in precipitation and smaller increases in temperature have increased runoff, although most cases have decreased runoff. Under the other methods, though, nearly all cases have increased runoff.

The changes in runoff propagate directly into changes in NBS (Fig. 1d), because the same results for the other components of NBS (overlake precipitation and evaporation) were used for all of the methods. Therefore, the discrepancies among methods are the same as for runoff. While the CC, PT, and EA methods generally show increased runoff, the balance of overlake precipitation and evaporation place their median changes in NBS at negative values. The details of this are beyond the scope of this study and vary among the GCMs and times of year, but as a general statement, overlake precipitation increases while overlake evaporation increases by even larger amounts.

The lake level of Lake Superior (Fig. 1e) is less sensitive to net basin supply than those of Lakes Michigan–Huron and Erie. This is in part because there is no other Great Lake upstream to enhance its sensitivity in NBS with an altered inflow, but more importantly because of the regulation of the outflow into the St. Marys River. One major goal in managing this regulated outflow is to maintain a near-constant lake level, thus prohibiting large changes in level. Nevertheless, projected lake levels under the TA method are markedly lower than under the other three methods. Although some data points lie above zero (rising lake level), the 75% quantile is below zero, so the great majority of GCM cases have falling lake levels under the TA method; the median is a drop of 24.5 cm. The CC, PT, and EA methods also have median changes in lake level that are negative, but of lesser magnitude (a drop of 1 cm under the PT method), and their 75% quantiles are all greater than zero, so they have significant chances of rises in lake level. Also, the results for the CC, PT, and EA methods all cluster much closer to one another than to the results of the TA method.

b. Lake Michigan–Huron basin

Lakes Michigan and Huron are considered a single unit because their attachment at the Straits of Mackinac is wide and deep enough for them to maintain the same hydraulic level. The results for Lakes Michigan–Huron, Erie, and Ontario largely reflect those of Lake Superior, although the results in terms of PET are less extremely dramatic, while the results in terms of lake levels are of greater magnitude (lake level results are not available for Lake Ontario).

The ratios of PET in the basin of Lake Michigan–Huron (Fig. 2a, land only) are a muted version of the results for Lake Superior (Fig. 1a), but again the changes in PET for the CC, PT, and EA methods cluster at values only moderately above unity; while under the TA method, the median value is just above 2, while several model runs project increases by factors greater than 4. These are not as extreme as the values for Lake Superior, mainly because the calibrated values of T_b are generally larger in the Michigan–Huron subbasins (falling mostly in the range of 5°–9°C) than those for
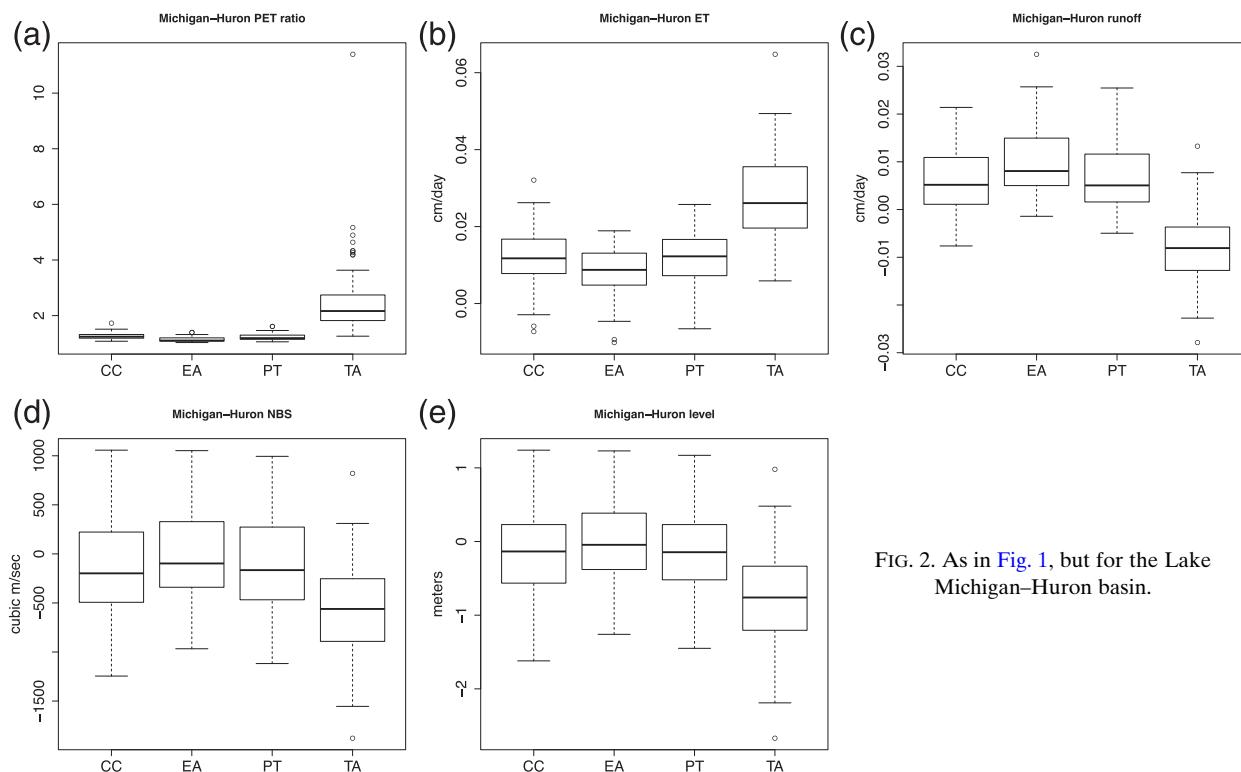
FIG. 2. As in Fig. 1, but for the Lake Michigan–Huron basin.

Superior; the reason for this is not clear and is part of the black-box nature of the calibration of LBRM. Nevertheless, these results are equivalent to solar input increased by two- to fourfold.

The actual ET from the Michigan–Huron basin (Fig. 2b, land only) is again modulated by availability of moisture. The increases in ET under the TA method strongly outstrip those under the other three methods, but those under the other three methods lean more strongly toward increased ET than in the Lake Superior basin (cf. Fig. 1b).

In terms of runoff in the Lake Michigan–Huron basin (Fig. 2c), as well as net basin supply (Fig. 2d), the results are qualitatively similar to Lake Superior and follow from the ET results. The runoff and NBS under the TA method are considerably lower than under the other three methods. As in Lake Superior, the CC, PT, and EA methods generally show increased runoff, but median decreases in NBS.

Lake levels on Lake Michigan–Huron are more sensitive than Lake Superior levels because of a combination of factors. In the model results from all methods, the changes in NBS are generally of the same sign for all lakes. This leads to the effect that Lake Superior's outflow due to its NBS has a reinforcing effect on the change in NBS in Lake Michigan–Huron. Furthermore, in contrast to Lake Superior, Lake

Michigan–Huron has no regulation at its outlet, so lake level variations cannot be artificially damped by direct human intervention. Although this can be changed through construction of regulation works in the future, we assume the status quo in this respect in order to isolate the effect due to climate. Finally, with the accumulation of changes in NBS from the Superior, Michigan–Huron, and Erie basins, the level of Lake Erie also changes in the same sense as the change in NBS for those combined basins. Because the difference in level between Lake Michigan–Huron and Lake Erie is small (just over 2 m), a significant rise in Lake Erie's level can inhibit outflow from Lake Michigan–Huron, requiring a rise in Michigan–Huron's level to restore dynamic equilibrium, and likewise for concurrent lake level drops on Lake Erie and Michigan–Huron. Thus, the change in lake levels for Lake Michigan–Huron (Fig. 2e) generally has the same sign as those for Lake Superior for an individual GCM realization, but the magnitudes are distinctly larger. Under the TA method, the 75% quantile is less than zero, indicating a large majority of GCM cases leading to decreased lake levels. The median value is a drop of 67.5 cm, and the largest drop is 267 cm in the IPSL RCP 8.5 run for the late twenty-first century. The TA results also have a large spread, with a few results showing rises in lake level around 1 m.
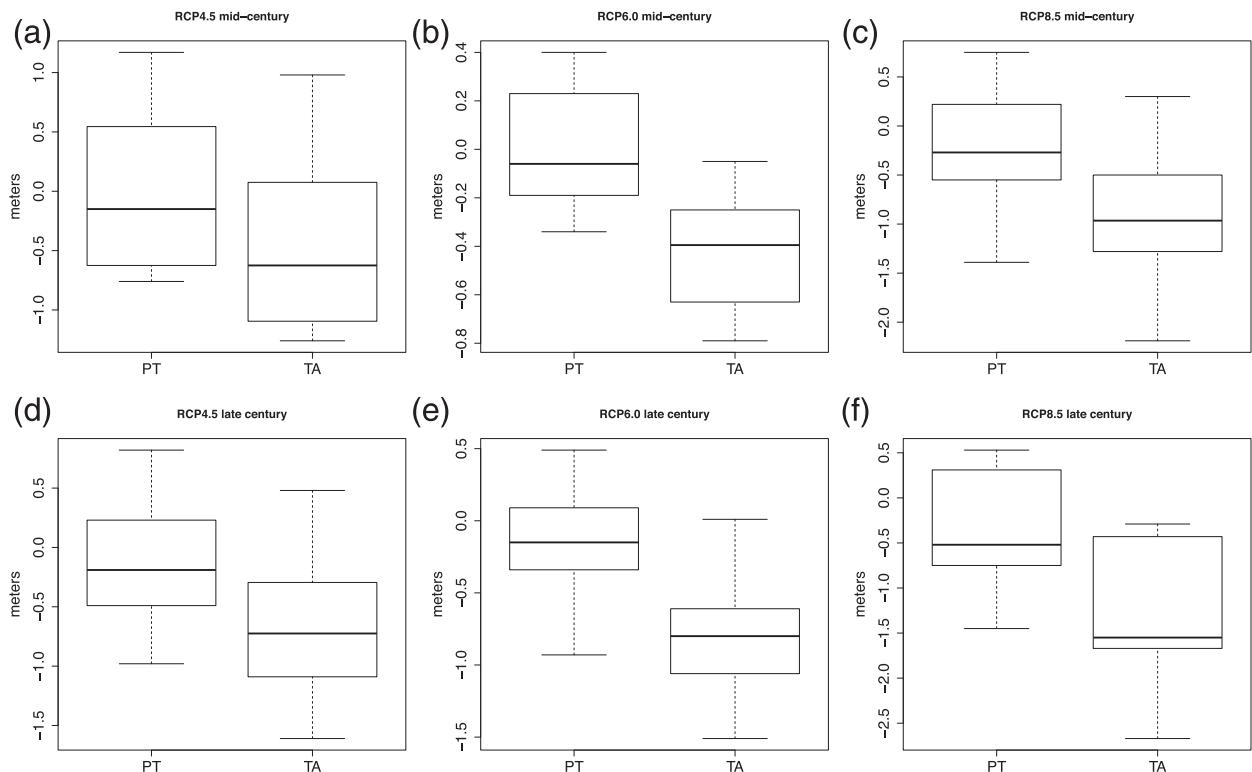
FIG. 3. Lake Michigan–Huron water levels represented using the TA and PT methods for subsets of the GCM realizations. Note that the scale differs among the panels. (a) Mid-twenty-first century for RCP 4.5, (b) mid-twenty-first century for RCP 6.0, (c) mid-twenty-first century for RCP 8.5, (d) late twenty-first century for RCP 4.5, (e) late twenty-first century for RCP 6.0, and (f) late twenty-first century for RCP 8.5. Each data point used in creating the plots represents an individual GCM realization.

The CC, PT, and EA results again cluster closer to one another than to the TA results, with median values less than zero and 75% quantile values greater than zero. The median lake level change under the PT method is a drop of 14.5 cm, 53.0 cm higher than the TA method, and the extreme low, again from the IPSL RCP 8.5 run for the late twenty-first century, is a drop of 145 cm, diminished by 122 cm relative to the extreme under the TA method. The PT and EA methods show a narrower range of changes between the 25% and 75% quantiles than the TA method.

To illustrate Lake Michigan–Huron's water level sensitivity to time horizon and RCP scenario, a comparison for subsets of the GCM realizations is given in Fig. 3. The sample sizes for these subsets are smaller—there are 12 GCM realizations contributing to the RCP 4.5 sets for each time period, and 10 each for RCP 6.0 and RCP 8.5. Figure 3 includes only the TA and PT methods, and for all of the combinations of RCP and time period, PT results in higher lake levels in all of the following measures: upper and lower extremes, 25% quantile, median, and 75% quantile. In Figs. 3b and 3e, the 25% quantile for the PT method is greater than the 75% quantile for the TA method, that is, the boxes do

not overlap. The median projected lake levels under the RCP 4.5 scenarios are comparable to or lower than those under the RCP 6.0 scenarios, in both the mid- and late century and using both the TA and PT methods, likely reflecting a different set of GCMs in the sample. The median lake level changes for all sampled groups are drops in lake level.

c. Lake Erie and Ontario basins

The results for simulations of future scenarios on Lakes Erie and Ontario are qualitatively similar to those for Lakes Superior (Fig. 1) and Michigan–Huron (Fig. 2). They are not shown here but are shown in Lofgren and Rouhana (2016). Note, however, that because of limitations in CGLRRM's ability to project Lake Ontario levels, those are not shown there, either.

d. Correlation of discrepancies with temperature change

The amount by which the TA method overestimates the influence of temperature on the various outputs is linked to the magnitude of the temperature change that is input. Figure 4 shows the discrepancy in Lake
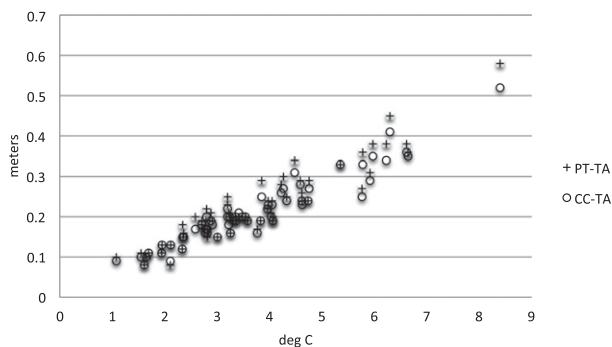
FIG. 4. Scatterplot of the difference in lake level (m) for Lake Superior, PT minus TA (plus signs) and CC minus TA (circles) vs change in air temperature. Each data point represents an individual GCM realization.

Superior lake level (result from the CC method minus the TA method and PT method minus TA method) as a function of the difference between the GCM-simulated historical and future air temperature. Lake Superior was chosen to eliminate the upstream influence from other lakes. Figure 4 shows very high correlations: $R = 0.951$ in the case of CC minus TA and $R = 0.932$ in the case of PT minus TA. This accounts for much of the narrowing of the distributions of all of the variables in Figs. 1–3 under the CC, PA, and EA methods relative to the TA method; the largest increases in PET and ET under the TA method are associated with the largest over-estimates and largest temperature increases, while the GCM cases with smaller temperature increases have smaller influences under the TA method and thus lesser disagreement with the other methods. The narrowing of the distributions under the other methods relative to TA is actually more distinct in the runoff and, especially, ET plots rather than the NBS and lake level plots.

e. Statistical significance

The discrepancies between the TA method and all three alternative methods are highly systematic, stemming first from the very large differences in the projected changes in PET. This leads to the results being biased similarly regardless of the GCM that is used as input. Therefore, the TA method's projected results are greater than those for the CC, PT, and EA methods in terms of PET and ET, and less in terms of runoff, NBS, and lake level. This result holds for all of the GCM realizations that we used for both of the time periods, and for all of the lakes.

One simple way to think of the statistical significance of these comparisons is to think of each comparison between the TA method and each alternative method as having equal chances of either result being greater or less than the other, under the null hypothesis that both

distributions are the same. However, this leaves us with the question of how many of these comparisons are independent of each other, for which there is not a clear answer. However, we will use a conservative guess at the number of independent samples that are present. Since PET controls ET, and in turn, runoff, NBS, and lake level, we will consider all of these variables to be dependent on each other and therefore together constitute only one independent variable. Likewise, we consider the separate results from each of the lakes to constitute only one independent variable, as well as the various results from different RCP scenarios and realizations of the same GCM, and the two different time periods that were compared. Thus, the only dimensions of independence that we have left are the comparison between the CC and TA method on the one hand and between the PT and TA method on the other hand (we regard PT and EA to not be independent of each other), as well as the eight different GCMs that were used. Thus, we regard the full set of experiments as resulting in 16 independent comparisons between the TA method and alternative methods, all of which show discrepancies of the same sign. The probability of this happening under the null hypothesis is $(0.5)^{16} = 1.53 \times 10^{-5}$ (single-tailed test), meaning that in terms of percentage significance, it is around 99.998%.

See Lofgren and Rouhana (2016) for the results of using the Wilcoxon rank sum test applied to historical and future lake levels across all of the model realizations, which yields even more extremely high estimates of statistical significance. These estimates of statistical significance are not perfectly rigorous because it is unclear how many of the realizations should be regarded as independent, but they indicate levels of statistical significance far beyond the often-used standard threshold of 95% or even 99%.

4. Discussion

The LBRM was carefully crafted to have features such as a mathematical solution that is entirely analytic within each time step (see Croley 1983) and therefore numerically stable under all conditions, as well as a calibration process that thoroughly searches the parameter space. It also conditions the energy available for evapotranspiration (which, divided by the factor of latent heat of vaporization per unit mass, we refer to as potential ET) on total insolation at the top of the atmosphere during the calibration runs. However, it seems that the question was not asked whether that energy constraint would carry over to other climatic regimes. One hint at the answer to this is how sensitive the LBRM's energy available for evapotranspiration is to

air temperature; the CC relationship of approximately a 7% $(°C)^{-1}$ increase is a significant benchmark as a value that should not be greatly exceeded.

However, LBRM's sensitivity is considerably greater. The values of T_b shown in Table 1 of Lofgren and Rouhana (2016) can be combined with air temperature changes of several degrees to verify the extreme changes in PET shown for the TA method in Figs. 1a and 2a. The most extreme case for a single subbasin is the combination of the Montreal River subbasin, at the eastern end of Lake Superior [listed in Lofgren and Rouhana (2016, Table 1) as Lake Superior subbasin 13], combined with the GFDL CM3 in the late twenty-first century under the RCP 8.5 scenario. In this case, the mean annual air temperature change is about 8.5°C, and the calibrated value of T_b is 1.1°C, so the factor of increase in PET is $e^{8.5°C/1.1°C} = 2269$. Other cases are less extreme but still highly significant. Because the base case's PET is constrained by Eq. (1) to be equivalent to the incident sunlight, these factors of increase in PET can be considered as factors of increase in equivalent incident sunlight, or, more vividly, as the equivalent number of suns in the sky of the virtual world created in the LBRM.

In LHW and here, caution has been used to say that in the TA method, air temperature is taken to be a proxy for PET that is equally valid across all time scales and climate regimes (the ''universal proxy'' assumption), and not to say that the TA method assumes that air temperature causes PET (the causation assumption). Now we will examine how these two assertions relate to one another. The causation assumption is a specific case of the universal proxy assumption. Functionally, in terms of setting up model simulations, these two types of assumption are entirely equivalent; the calculations that are used in the modeling are the same whether one regards air temperature as a universal proxy for PET or as an actual cause. However, as demonstrated in LHW by the contrast in air temperature–ET correlation on the basis of time scale, the relation between these variables is not universal, but highly dependent on the time scale examined (assertion 2 from the introduction). In reality, the annual cycle in air temperature is driven by net radiation (led by incoming solar radiation), which strongly drives ET, while temperature variability on other time scales is driven primarily by variability in heat advection, which drives ET less strongly. Warming by greenhouse gases is yet another mechanism for driving air temperatures, which can be expected to have its own mode of mapping onto ET.

In Croley (1990), it is stated that the change function method was used at the instruction of the leaders of the parent project, an evaluation of overall impacts of climate change funded by the U.S. Environmental Protection Agency (Smith and Tirpak 1989). But the change function method, which assumes that the user will apply it to the proper variables as drivers, was developed separately from the LBRM, which assumes a universal proxy relationship for PET based on air temperature. Consideration of the interaction of these assumptions seems not to have been strongly considered at that time.

Some other methods are popular for assessment of climate change's impact on water resources and ET, using a one-way coupling of data feeding from a GCM to a hydrologic model. One of the most popular uses the PET equation of Thornthwaite (1948), often within the context of the Palmer drought severity index. McAfee (2013) uses the PET formulation of Hamon (1961) as a point of reference. Thornthwaite (1948), Hamon (1961), and similar formulations use aspects of the seasonal cycle, often in terms of length of daylight, as explicit predictors within their formulation. This contrasts with LBRM, in which all of the variability in PET is attributed to air temperature as the predictor (regardless of whether this is framed as the universal proxy assumption or the causation assumption). By attributing much of the variability within historical times directly to the seasonal cycle, instead of to air temperature, the sensitivity to air temperature under these other formulations is much less than in LBRM. The Hamon (1961) formulation uses saturation vapor pressure as the main temperature-dependent term, making it very similar to our CC method. This method, which we regarded as an outer bound for PET sensitivity and has been shown to be much less sensitive to air temperature than LBRM, is stated to be excessively temperature sensitive in Shaw and Riha (2011) and McAfee (2013).

Weart (2015) calls climate change impacts research ''a peculiar kind of science'' because it covers a wide range of end goals for prediction, and even for a single type of impact, multiple methods can exist side by side with no clear correct approach. The results from various studies are then regularly compiled into synthesis documents by committees of scientists. This paper and its predecessor, LHW, attempt to show an example of how arguments based on fundamental physical constraints can be used to distinguish reasonable from unreasonable methods of climate change impact assessment.

Because of smaller alterations in air temperature, applications of LBRM to subseasonal to interannual time scales are likely to have much smaller issues with its formulation (see Gronewold et al. 2011; Fry et al. 2014). Additional assessment of LBRM in the context of the connection between air temperature anomalies and errors in runoff in historical periods is warranted.

5. Conclusions and future plans

A method for projecting the water levels of the Laurentian Great Lakes under scenarios of human-caused climate change, used almost to the exclusion of other methods in the past, relies very heavily on the large basin runoff model (LBRM) as a component for determining the water budget for the lake system. LBRM's exclusive reliance on air temperature as a predictor of PET and its very high sensitivity to air temperature cause it to overestimate future ET and cause the entire modeling system to underestimate future lake levels. This excessive sensitivity is at least partially due to LBRM's failure to distinguish between the ET–air temperature correlation associated with the annual cycle (therefore driven largely by changes in incoming solar radiation and net surface radiation) and variability in ET and air temperature associated with other time scales and other causes.

We created three alternative methods for transferring climate outputs from GCMs into the LBRM: energy adjustment (EA), Priestley–Taylor (PT), and Clausius–Clapeyron (CC). We deem all of these alternative methods, especially the PT method, to be more reflective of what is actually happening within the GCM than the TA method. All three of these alternative methods show, relative to the temperature adjustment method, less PET and ET, more runoff from the land and net basin supply for the lake basins, and higher lake water levels in the future. The results from the EA, PT, and CC methods cluster much closer to one another than to the TA method results. The median change in lake level of Lake Michigan–Huron among all the GCM runs shifts from a drop of 67.5 cm under the TA method to a drop of only 14.5 cm under the PT method. This contrast puts the PT method's median estimate of lake level change well within the range of historical variability.

The magnitude of the discrepancies between the TA and other methods is highly correlated with the air temperature change in the driving GCM (larger temperature changes lead to larger discrepancies; Fig. 4). This narrows the distributions of changes in all of the variables among the driving GCM runs, although the spread still remains sizable.

Using various methods of estimating the statistical significance, we find that, at minimum, these discrepancies in results are significant at the 99.998% level. Because of the highly systematic sources of error within the formulation of the LBRM under the TA method, we regard these extremely high values of statistical significance as inevitable and largely beside the point—simply the result of a model formulation that, intentionally or not, yields excessive sensitivity of ET to air temperature.

The results here apply only to the water levels of the Great Lakes and the hydrology of its basin and do not affect larger questions of climate change. In fact, the alternative methods for hydrologic sensitivity used here present a truer picture of what the GCMs are telling us about climate change's impacts, as they are using information from them to more properly drive the surface energy budget and hydrology. However, there are lessons for the broader impacts community:

1) When examining impacts, be sure to distinguish between effects of radiation that can be modulated by season or latitude and impacts that are directly driven by air temperature.
2) Understand that climate change is not merely a group of physical quantities output by models but is a set of interacting processes, and that two-way surface–atmosphere interactions are crucial in creating climate change.
3) Use the greatest of caution if you are going to significantly second-guess the climate model results that relate directly to your impact of interest (ET and surface net radiation, in our case).

The alternative methods of transferring GCM-predicted variables into the LBRM for future projections are relatively quick and easy modifications to the existing modeling system, for purposes of demonstration of the problems with the previously used method, and are not final answers. As a next step, therefore, we would like to put forward a larger effort to model the Great Lakes basin's water budget using a hydrologic model with a strong basis in the surface energy budget and driven by a variety of GCM data using a more modern climate downscaling technique.

REFERENCES

Angel, J. R., and K. E. Kunkel, 2010: The response of Great Lakes water levels to future climate scenarios with an emphasis on Lake Michigan–Huron. *J. Great Lakes Res.*, **36**, 51–58, doi:10.1016/j.jglr.2009.09.006.

Brean, H., 2015: Climate change will increase evaporation of Colorado River. *Las Vegas Review-Journal*, 6 February, accessed 12 May 2015. [Available online at http://www.reviewjournal.com/news/water-environment/climate-change-will-increase-evaporation-colorado-river.]

Chao, P., 1999: Great Lakes water resources: Climate change impact analysis with transient GCM scenarios. *J. Amer. Water Resour. Assoc.*, **35**, 1499–1507, doi:10.1111/j.1752-1688.1999.tb04233.x.

Collins, W. J., and Coauthors, 2011: Development and evaluation of an Earth-system model—HadGEM2. *Geosci. Model Dev.*, **4**, 1051–1075, doi:10.5194/gmd-4-1051-2011.

Croley, T. E., II, 1983: Great Lakes basins (U.S.A.–Canada) runoff modeling. *J. Hydrol.*, **64**, 135–158, doi:10.1016/0022-1694(83)90065-3.

——, 1989: Verifiable evaporation modeling on the Laurentian Great Lakes. *Water Resour. Res.*, **25**, 781–792, doi:10.1029/WR025i005p00781.

——, 1990: Laurentian Great Lakes double-CO_2 climate change hydrological impacts. *Climatic Change*, **17**, 27–47, doi:10.1007/BF00148999.

——, and H. C. Hartmann, 1984: Lake Superior basin runoff modeling. NOAA Tech. Memo. ERL GLERL-50, 284 pp. [Available online at http://www.glerl.noaa.gov/ftp/publications/tech_reports/glerl-050/tm-050.pdf.]

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi:10.1175/JCLI3629.1.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson, 1986: Biosphere–Atmosphere Transfer Scheme (BATS) for the Community Climate Model. NCAR Tech. Note NCAR/TN-275+STR, 72 pp., doi:10.5065/D6668B58.

Donner, L. J., and Coauthors, 2011: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. *J. Climate*, **24**, 3484–3519, doi:10.1175/2011JCLI3955.1.

d'Orgeville, M., W. R. Peltier, A. R. Erler, and J. Gula, 2014: Climate change impacts on Great Lakes basin precipitation extremes. *J. Geophys. Res. Atmos.*, **119**, 10 799–10 812, doi:10.1002/2014JD021855.

Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, doi:10.1007/s00382-012-1636-1.

Dunne, J. P., and Coauthors, 2012: GFDL's ESM2 global coupled climate–carbon Earth system models. Part I: Physical formulation and baseline simulation characteristics. *J. Climate*, **25**, 6646–6665, doi:10.1175/JCLI-D-11-00560.1.

——, and Coauthors, 2013: GFDL's ESM2 global coupled climate–carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate*, **26**, 2247–2267, doi:10.1175/JCLI-D-12-00150.1.

Fry, L. M., and Coauthors, 2014: The Great Lakes Runoff Intercomparison Project Phase 1: Lake Michigan (GRIP-M). *J. Hydrol.*, **519**, 3448–3465, doi:10.1016/j.jhydrol.2014.07.021.

Government of Canada and U.S. EPA, 1995: *The Great Lakes: An Environmental Atlas and Resource Book*. 3rd ed. Tech. Rep. EPA 905-B-95-001, 46 pp.

Great Lakes Regional Assessment Group, 2000: Preparing for a changing climate: The potential consequences of climate variability and change—Great Lakes overview. P. Sousounis and J. Bisanz, Eds., Rep. for the U.S. Global Change Research Program, University of Michigan, 106 pp. [Available online at http://geo.msu.edu/extra/glra/PDF_files/GLRA_report.pdf.]

Gronewold, A. D., A. H. Clites, T. S. Hunter, and C. A. Stow, 2011: An appraisal of the Great Lakes Advanced Hydrologic Prediction System. *J. Great Lakes Res.*, **37**, 577–583, doi:10.1016/j.jglr.2011.06.010.

Hamon, W. R., 1961: Estimating potential evapotranspiration. *Proc. Amer. Soc. Civ. Eng.*, **87**, 107–120.

Hartmann, H. C., 1990: Climate change impacts on Laurentian Great Lakes levels. *Climatic Change*, **17**, 49–67, doi:10.1007/BF00149000.

Hayhoe, K., J. VanDorn, T. Croley II, N. Schlegal, and D. Wuebbles, 2010: Regional climate change projections for Chicago and the Great Lakes. *J. Great Lakes Res.*, **36**, 7–21, doi:10.1016/j.jglr.2010.03.012.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, doi:10.1175/JCLI3990.1.

Hobbins, M. T., A. Dai, M. L. Roderick, and G. D. Farquhar, 2008: Revisiting the parameterization of potential evaporation as a driver of long-term water balance trends. *Geophys. Res. Lett.*, **35**, L12403, doi:10.1029/2008GL033840.

Hurrell, J., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, doi:10.1175/BAMS-D-12-00121.1.

International Lake Ontario–St. Lawrence River Study Board, 2006: Options for managing Lake Ontario and St. Lawrence River water levels and flows. Final Rep. to the International Joint Commission, 262 pp. [Available online at http://www.ijc.org/loslr/en/library/LOSLR%20Study%20Reports/report-main-e-6KB.pdf.]

Kling, G. W., and Coauthors, 2003: Confronting climate change in the Great Lakes Region: Impacts on our communities and ecosystems. Rep. of the Union of Concerned Scientists and Ecological Society of America, 92 pp. [Available online at http://www.ucsusa.org/sites/default/files/legacy/assets/documents/global_warming/greatlakes_final.pdf.]

Kutzbach, J. E., J. W. Williams, and S. J. Vavrus, 2005: Simulated 21st century changes in regional water balance of the Great Lakes region and links to changes in global temperature and poleward moisture transport. *Geophys. Res. Lett.*, **32**, L17707, doi:10.1029/2005GL023506.

Lide, D. R., Ed., 2005: *CRC Handbook of Chemistry and Physics*. 86th ed. CRC Press, 2712 pp.

Lofgren, B. M., 2014: Simulation of atmospheric and lake conditions of the Laurentian Great Lakes region using the Coupled Hydrosphere–Atmosphere Research Model (CHARM). NOAA Tech. Memo. GLERL-165, 23 pp. [Available online at http://www.glerl.noaa.gov/ftp/publications/tech_reports/glerl-165/tm-165.pdf.]

——, and J. Rouhana, 2016: Reaffirmation of large biases in a long-used method for projecting changes in Great Lakes water levels in climate change scenarios. NOAA Tech. Memo. GLERL-167, 29 pp. [Available online at http://www.glerl.noaa.gov/ftp/publications/tech_reports/glerl-167/tm-167.pdf.]

——, F. H. Quinn, A. H. Clites, R. A. Assel, A. J. Eberhardt, and C. L. Luukkonen, 2002: Evaluation of potential impacts on Great Lakes water resources based on climate scenarios of two GCMs. *J. Great Lakes Res.*, **28**, 537–554, doi:10.1016/S0380-1330(02)70604-7.

——, T. A. Hunter, and J. Wilbarger, 2011: Effects of using air temperature as a proxy for potential evapotranspiration in climate change scenarios of Great Lakes basin hydrology. *J. Great Lakes Res.*, **37**, 744–752, doi:10.1016/j.jglr.2011.09.006.

Lorenz, D. J., and E. T. DeWeaver, 2007: The response of the extratropical hydrological cycle to global warming. *J. Climate*, **20**, 3470–3484, doi:10.1175/JCLI4192.1.

MacKay, M., and F. Seglenieks, 2013: On the simulation of Laurentian Great Lakes water levels under projections of

global climate change. *Climatic Change*, **117**, 55–67, doi:10.1007/s10584-012-0560-z.

Manabe, S., and J. L. Holloway, 1975: The seasonal variation of the hydrologic cycle as simulated by a global model of the atmosphere. *J. Geophys. Res.*, **80**, 1617–1649, doi:10.1029/JC080i012p01617.

——, R. T. Wetherald, P. C. D. Milly, T. L. Delworth, and R. J. Stouffer, 2004: Century-scale change in water availability: CO_2-quadrupling experiment. *Climatic Change*, **64**, 59–76, doi:10.1023/B:CLIM.0000024674.37725.ca.

Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geosci. Model Dev.*, **4**, 723–757, doi:10.5194/gmd-4-723-2011.

McAfee, S. A., 2013: Methodological differences in projected potential evapotranspiration. *Climatic Change*, **120**, 915–930, doi:10.1007/s10584-013-0864-7.

Milly, P. C. D., K. A. Dunne, and A. V. Vecchia, 2005: Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, **438**, 347–350, doi:10.1038/nature04312.

Monteith, J. L., 1973: *Principles of Environmental Physics*. Edward Arnold, 241 pp.

Morton, F. I., 1983: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *J. Hydrol.*, **66**, 1–76, doi:10.1016/0022-1694(83)90177-4.

Notaro, M., V. Bennington, and B. Lofgren, 2015: Dynamical downscaling–based projections of Great Lakes water levels. *J. Climate*, **28**, 9721–9745, doi:10.1175/JCLI-D-14-00847.1.

Priestley, C. H. B., and R. J. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.

Pryor, S. C., D. Scavia, C. Downer, M. Gaden, L. Iverson, R. Nordstrom, J. Patz, and G. P. Robertson, 2014: Midwest. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, T. C. Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 418–440, doi:10.7930/J0KS6PHH.

Rotstayn, L. D., S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. D. Hirst, J. I. Syktus, and K. K. Wong, 2012: Aerosol- and greenhouse gas–induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations. *Atmos. Chem. Phys.*, **12**, 6377–6404, doi:10.5194/acp-12-6377-2012.

Scheff, J., and D. M. W. Frierson, 2014: Scaling potential evapotranspiration with greenhouse warming. *J. Climate*, **27**, 1539–1558, doi:10.1175/JCLI-D-13-00233.1.

Shaw, S. B., and S. J. Riha, 2011: Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests. *Hydrol. Processes*, **25**, 1466–1478, doi:10.1002/hyp.7913.

Smith, J. B., and D. Tirpak, Eds., 1989: The potential effects of global climate change on the United States. Rep. EPA-230-05-89-050, U.S. EPA, 413 pp.

Thornthwaite, C. W., 1948: An approach toward a rational classification of climate. *Geogr. Rev.*, **38**, 55–94, doi:10.2307/210739.

Tolson, B., 2009: Rationalizing the decline in Lake Michigan–Huron water levels using the Coordinated Great Lakes Routing and Regulation Model. Final Rep., 35 pp. [Available at http://iugls.org/DocStore/ProjectArchive/HCL_Hydroclimate/HCL08_Tolson_RationalizingDeclineMHLevelsWithCGLRRM/Reports/HCL08-R1_Tolson.pdf.]

Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, doi:10.1007/s00382-011-1259-y.

Walsh, J., and Coauthors, 2014: Appendix 3: Climate Science Supplement. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, T. C. Richmond, G. W. Yohe, Eds., U.S. Global Change Research Program, 735–789, doi:10.7930/J0KS6PHH.

Weart, S., 2015: Climate change impacts: The growth of understanding. *Phys. Today*, **68**, 46–52, doi:10.1063/PT.3.2914.