Journal of Hydrology 601 (2021) 126584

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Modeling the impact of spatiotemporal vegetation dynamics on groundwater recharge

Harsh Anurag [a,*], G.H. Crystal Ng [a,b], Robert Tipping [c], Kathy Tokos [d]

[a] Department of Earth & Environmental Sciences, University of Minnesota – Twin Cities, Minneapolis, MN 55455, USA
[b] Saint Anthony Falls Laboratory, University of Minnesota – Twin Cities, Minneapolis, MN 55414, USA
[c] Minnesota Department of Health, St. Paul, MN 55164, USA
[d] Formerly at Department of Earth & Environmental Sciences, University of Minnesota – Twin Cities, Minneapolis, MN 55455, USA

ARTICLE INFO

Keywords:
Groundwater recharge
Dynamic vegetation
Data assimilation
Ensemble Kalman filter
Ecohydrology

ABSTRACT

Climate change affects the growth of vegetation and its physiological states such as leaf area index (LAI), which in turn can affect groundwater recharge because of changes in evapotranspiration (ET). Presently, most recharge modeling studies over-simplify transient vegetation conditions and the potential corresponding impact on recharge by using climatological values of vegetation parameters such as LAI. Our study uses the Community Land Model (CLMv4.5) to investigate the sensitivity of recharge to interannual varying vegetation in Minnesota (USA) across different climate, hydrogeology, and ecoregions at a 25 km spatial resolution and for the period of 2000–2015. The Ensemble Kalman Filter (EnKF) was used to calibrate soil and runoff parameters to statewide water table depth observations. Results of the study indicate that although year-to-year varying vegetation does not affect long-term climatological recharge estimates, it can drive disproportionately large variability in annual and seasonal recharge. Comparing simulations with dynamic and climatological vegetation inputs, the average magnitude difference (root mean square difference, RMSD) for recharge was 21.1% in response to only a 4% difference in LAI inputs. Regression analysis revealed that the combination of local hydrogeology and vegetation-type affects the magnitude of recharge response to LAI and ET changes. We also found cross-ecoregion dominance of temperature rather than precipitation controlling LAI anomalies and resulting recharge variability, with springtime temperature being the primary factor because of its impact on leaf-out conditions. Drier western Minnesota showed higher relative LAI differences as well as higher spring time and relative annual recharge compared to the wetter eastern part of the state, indicating higher vulnerability of the water-limited region to changing vegetation and climatic conditions. Our study shows that models can underestimate or overestimate annual and seasonal recharge if vegetation dynamics are neglected, demonstrating the need to incorporate transient vegetation conditions when assessing the impact of future climate change on recharge.

1. Introduction

Vegetation interacts with the water cycle in multiple ways (Rodriguez-Iturbe, 2000). Leaves intercept precipitation before it either evaporates from the canopy surface or moves to the ground and infiltrates. Plant litter also alters the water retention and infiltration properties of the soil. However, the most important link in terms of water balance is usually the effect of vegetation on soil moisture through transpiration (Arora, 2002). Through these processes, vegetation directly influences the amount of water that percolates past the root zone and recharges the water table. In particular, plant physiological properties such as amount of leaves and root distribution control the amount of water being intercepted and/or transpired (Le Maitre et al., 1999). Thus, any changes in these plant physiological properties have the potential to induce changes in hydrological processes, including groundwater recharge.

In hydrological or land-surface models, vegetation parameters usually include leaf area index (LAI), which is the ratio of the one-sided leaf area to the underlying ground area. LAI serves as a key functional plant trait because it controls the magnitude of important ecohydrological fluxes such as transpiration, canopy evaporation, and photosynthesis (Asner et al., 2003; Zheng and Moskal, 2009; Bondeau et al., 1999). As such, most models calculate these fluxes on the gridcell scale through scaling by LAI values. LAI is commonly specified in the model using

* Corresponding author.
E-mail address: anura003@umn.edu (H. Anurag).

https://doi.org/10.1016/j.jhydrol.2021.126584
Received 28 February 2021; Received in revised form 12 June 2021; Accepted 15 June 2021
Available online 18 June 2021

repeated climatological monthly values. This representation of LAI neglects the response of vegetation to interannual variability in climatic factors (Arora, 2002; Fatichi et al., 2015). In reality, changes in environmental conditions such as shifts in temperature and precipitation may influence the growth of vegetation (Jeong et al., 2011; Gunderson et al., 2012). For example, an earlier or a warmer spring might lead to an early bud burst in plants, while an extended period of moisture stress may hinder growth.

Seasonal and annual changes in vegetation in turn have the potential to affect hydrologic processes at various temporal or spatial scales. There have been previous studies that have investigated the impact of dynamic vegetation on different components of the water cycle apart from recharge using physically based hydrologic or land-surface models. With a focus on runoff, Zhang et al. (2009) and Tesemma (2015) included transient LAI in their model simulations and found significant improvement in the runoff estimate when compared to model simulations with climatological vegetation. Looking at the impact on evapotranspiration (ET) and soil moisture, Tang et al. (2012) and Vivoni (2012) found that substantial year-to-year variability in vegetation greening onset and dormancy dates generate significant differences in ET simulations compared to results using climatological LAI. Ford and Quiring (2013) showed that adding dynamic LAI improved simulated soil moisture and predictions of the intensity and duration of drought conditions. Hwang et al. (2018) analyzed precipitation-runoff deficits (i.e., the difference between precipitation and runoff) and concluded that the long-term non-stationary hydrologic behavior could not be simulated unless full vegetation dynamics, including vegetation phenology and long-term growth, were incorporated into the model. These studies provide strong evidence of the influence of dynamic vegetation on important hydrologic components (ET, runoff, soil moisture) and prompts questions whether similar interactions between vegetation and recharge occur.

Very few studies have specifically addressed the influence of dynamic LAI on groundwater recharge (Wegehenkel, 2009), despite the importance of recharge in determining groundwater resource availability. However, previous work has established the general sensitivity of groundwater to vegetation. Although early seminal work in ecohydrology (Eagleson, 1978; Rodriguez-Iturbe et al., 1999; Porporato et al., 2001; Fernandez-Illescas et al., 2001) did not explicitly consider this connection in their focus on dryland regions, where deep water tables are largely disconnected from root zone soil moisture, tight coupling has long been recognized in humid lands and wetlands in field-based observations (Peck and Williamson, 1987; Chang, 2012), and more recently in model-based studies (Laio et al., 2009; Loheide and Gorelick, 2007). The influence of vegetation on recharge has been more widely demonstrated across diverse environments across the globe through studies investigating land use/land cover effects (LU/LC) on recharge, including both spatial variability and temporal shifts in LU/LC (Scanlon et al., 2006; Petheram et al., 2002; Kim and Jackson, 2012). These LU/LC studies, however, focused on changes in vegetation type and did not consider the dynamic growth variations within a vegetation type.

The impact of more subtle vegetation growth variations on groundwater recharge, without vegetation type change, has received less attention. This serves a critical knowledge gap, because climate change is predicted to alter plant phenology and leaf carbon stocks (Xu et al., 2013; Loheide and Gorelick, 2013; Zhu et al., 2016; Wu et al., 2015), and these changes directly impact ET and subsequently influence the amount of water that is available to recharge the water table. Most work on the impact of climate change on recharge included yearly repeated (climatological) inputs for vegetation parameters (e.g. Jyrkama and Sykes, 2007; Döll, 2009; Ng et al., 2010; Hanson and Dettinger, 2005; Bouraoui et al., 1999; Herrera-Pantoja and Hiscock, 2007; Chen et al., 2002; Croley and Luukkonen, 2003; Kirshen, 2002) and thus fail to account for vegetation responses that can drive feedbacks among climate, vegetation, and water (Taylor et al., 2012; Mao et al., 2016). A smaller number of climate change studies (Eckhardt and Ulbrich, 2003; Green et al., 2007; Crosbie et al., 2010; Wegehenkel, 2009) have included dynamic representations of vegetation, including transient LAI, in their predictions of recharge. Although these recent studies reflect growing recognition of vegetation dynamics, none, with the exception of Wegehenkel (2009) conducted an assessment of the impact of dynamic vegetation on recharge predictions. In their analysis of a catchment in northeastern Germany over 1950–2100, Wegehenkel (2009) first executed two reference simulations using climatological and dynamic vegetation for 1950–1966 and then compared those with future projections over 2078–2094, again with the two vegetation scenarios. They found that cumulative groundwater recharge showed a decrease of 9% from the reference to the future period when using dynamic vegetation, compared to a decrease of 22% when using climatological vegetation. They attributed this difference to feedbacks among transpiration, soil water availability, and crop growth in the case with dynamic vegetation. Findings from this case study motivates further work to quantify the effect of climatically driven vegetation dynamics on recharge across climates and plant types. A fuller understanding of the interactions among climate, vegetation, and groundwater can help to determine the complexity needed for representing vegetation in future climate change assessments of groundwater resources.

Obstacles to understanding the influence of dynamic vegetation on recharge include uncertainties in constraining their interactions outside of regions where roots are directly in contact with groundwater (e.g., outside of wetlands and riparian zones). Although recent advances have led to more mechanistic representation of processes spanning the canopy to the deeper subsurface (Clark et al., 2015), challenges remain in calibrating such models to reliably simulate processes across the entire soil column, where soil properties are typically uncertain. Our study tackles this obstacle by implementing an efficient data assimilation method to calibrate regional land-surface model simulations of recharge, which enables us to address knowledge gaps about the influence of dynamic LAI on groundwater recharge. This modeling work is carried out using dynamic remotely sensed LAI observations across the state of Minnesota (USA). The state's precipitation and temperature gradients, and its varied ecoregions that include prairie, cropland, deciduous forest, evergreen forest, and mixed forest provide an apt testbed to evaluate the impact of spatiotemporal variability in vegetation on groundwater recharge across a range of conditions. Specifically, we address the following questions, using Minnesota as a test case: (i) How has groundwater recharge, over seasonal to multi-year time scales, responded to interannual changes in LAI in different ecoregions? (ii) What has been the relative importance of different climatic factors in driving interannual vegetation dynamics that influence groundwater recharge? Answers to these questions can help guide future predictions about climate change impacts on groundwater recharge across diverse continental ecoregions.

2. Methods

2.1. Study area

Minnesota experiences a continental climate with hot summers and cold winters. It has a distinct precipitation gradient with an average annual precipitation of about 89 cm in the southeast that gradually reduces to about 50 cm in the northwest (Fig. 1a). About two-thirds of the precipitation in the state occurs during May to September. The US Environmental Pollution Agency (EPA) divides Minnesota into five major ecological provinces, or ecoregions (Fig. 1b) (White, 2020). We will refer to the first of these ecoregions as Great Plains (GP) in the west and southwestern part of the state. These were once covered with tallgrass prairie, but now over 75% of the area is cropland. We further subdivided GP into Northern Great Plains (NGP) and Southern Great Plains (SGP) on the basis of soil texture. Both NGP and SGP are composed of glacial till, but glacial Lake Agassiz in the last glacial period deposited thick beds of lake sediment on top of glacial till in the north.

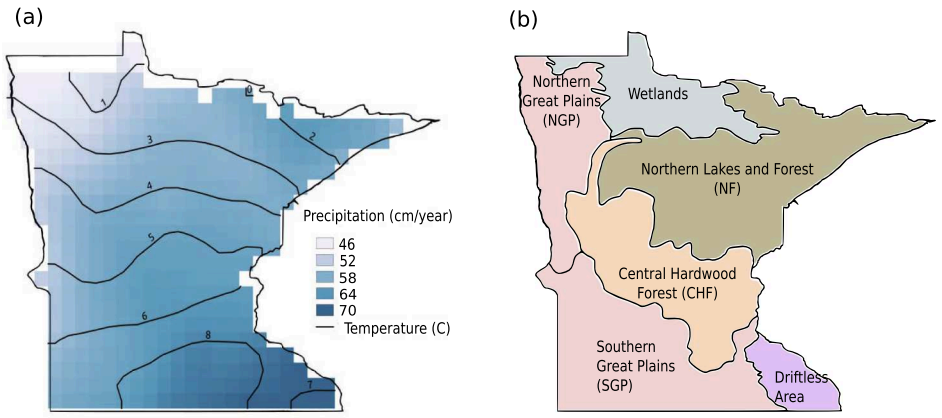

Fig. 1. (a) 16-year average (2000–2015) precipitation and temperature in Minnesota (from Thornton et al., 2014) (b) Major ecoregions in Minnesota. Our analysis focuses on the Northern and Southern Great Plains, Central Hardwood Forest, and Northern Forest (see Table 1), due to large uncertainties in simulating hydrologic conditions in the Wetlands and karst-dominated Driftless Area.

NGP is also distinct with its lower crop cover and smaller-statured boreal forests. The Central Hardwood Forests (CHF) ecoregion, extending from the southeastern to northwestern part of Minnesota, consists of temperate deciduous forests along with extensive cropland, pasture, and dairy operations. The Northern Forests (NF) ecoregion, occupying the northern and northeastern part of the state, is a region of mixed coniferous and deciduous temperate forests, conifer bogs, and broad lacustrine basins. The Northern Wetlands province consist of vast stretches of peat soil on former lake beds with some forest and several eroded river channels. The Driftless Area province in the southeast is a karst-dominated region characterized by forested ridges and carved river valleys. Because of the limitations of common hydrological and land-surface models in reliably simulating groundwater recharge in the complex systems of wetlands and karst regions (Ringeval et al., 2012; Hartmann et al., 2017), we focus our ecoregion-specific analyses on the four regions of NGP, SGP, CHF, and NF (key characteristics summarized in Table 1; list of abbreviations in Table 4).

2.2. Model description

We used the Community Land Model (CLM) v4.5 (Oleson et al., 2013) for our assessment of dynamic vegetation controls on groundwater recharge. CLM is a spatially distributed, mechanistic land-surface model that serves as the terrestrial component of the National Center for Atmospheric Research's (NCAR) larger Community Earth System Model (CESM) (Hurrell et al., 2013). CLMv4.5 has process-based representations of land-surface biogeophysics, biogeochemistry, hydrology, and vegetation (Oleson et al., 2013). Because its domain spans the land surface to unconfined aquifers, CLM is suitable for evaluating climate and vegetation controls on recharge to surficial aquifers. CLM lacks representation of deeper subsurface units and applies an approximate solution for lateral subsurface flow (described below). We consider the omission of a more detailed lateral subsurface flow scheme to be an acceptable approximation considering our target modeling resolution of 25 km for this regional scale application. Accurate simulation of lateral groundwater flow requires a higher spatial resolution due to its dependence on gravity and local topographic factors (Zeng et al., 2018). Moreover, Minnesota has a relatively low-relief topography (minimum of 183 m.a.s.l. and maximum of 701 m.a.s.l. over 225,181 km^2) that likely reduces the consequences of simplifying lateral groundwater movement.

The default representation of vegetation in CLM uses satellite-based monthly climatological LAI obtained from gridded datasets. We refer to this climatological satellite vegetation configuration as CLM_{CS}. Over the last decade, several land-surface models including CLM have added process-based modules to simulate plant carbon and nitrogen stocks, as well as LAI, as prognostic variables. There has also been progress in developing Dynamic Global Vegetation Models (DGVM) within land-surface models that can predict shifts in vegetation types over space and time (i.e. transient PFTs) (e.g. Druel et al., 2017; Clark et al., 2011;

Table 1
Key characteristics of different ecoregions.

Ecoregions	Terrain Characteristics	Major Vegetation (% land cover of the ecoregion)
Northern Great Plains (NGP)	Thick lacustrine sediment underlain by glacial till	• Crop (49%) • Grassland (34%) • Broadleaf Deciduous Boreal Forest (6%)
Southern Great Plains (SGP)	Homogeneous topography of level to gently rolling glacial till plains with areas of morainal hills and loess deposits	• Crop (71%) • Grassland (24%) • Broadleaf Deciduous Temperate Forest (3%)
Central Hardwood Forest (CHF)	Nearly level to rolling till plains, lacustrine basins, outwash plains, and moraines. Contains Minneapolis and St. Paul metropolis.	• Crop (45%) • Grassland (36%) • Broadleaf Deciduous Temperate Forest (8%)
Northern Forests (NF)	Nutrient-poor glacial soils, undulating till plains and broad lacustrine basins	• Broadleaf Deciduous Temperate Forest (25%) • Needleleaf Evergreen Temperate Forest (24%) • Crop (16%) • Grassland (16%)

Levis et al., 2004; Fisher et al., 2017). Although CLM incorporates optional modes for both these capabilities (CLM-Biogeochemical[CLM-BGC] and CLM-DGVM), we instead modified the general CLM code to directly take as inputs spatiotemporally dynamic LAI from satellite observations. This strategy has the advantage over the BGC mode of avoiding uncertainties in prognostic plant variable simulations, which depend on a large number of vegetation parameters that substantially add to the computational load for model calibration.

Because the model integrates over multiple plant-functional-types (PFTs) within a grid cell to simulate total transpiration (and other vegetation fluxes), we developed a scheme to convert satellite-observed total LAI into PFT-level LAI based on the disaggregation in the default climatological CLM dataset:

$$\frac{LAI_{i,SAT}}{TLAI_{SAT}} = \frac{LAI_{i,CLM\_DEF}}{TLAI_{CLM\_DEF}}, \tag{1}$$

where LAI_i and $TLAI$ are PFT-level LAI (for the ith PFT out of a total of 15 PFTs represented in CLM) and total LAI respectively, *SAT* denotes LAI observed from satellite, and *CLM_DEF* indicates the default CLM dataset that is provided with the model. Our scheme assumes that the default CLM spatial distributions of PFT cover applies for our disaggregation calculation; ecoregion averages of percent cover of each PFT are included in Table 1 (full spatial distribution shown in Fig. S9 in Supplementary Information). CLM's default climatological LAI ($TLAI_{CLM\_DEF}$) and PFT cover dataset is a spatially distributed product at a base resolution of 0.05° and is derived from MODIS satellite data over the years 2001–2003 (Lawrence and Chase, 2007). This time period for the default CLM dataset overlaps with the early part of our simulation period (2000–2015), and we consider the default CLM PFT distribution to reasonably apply over later years as well, given that more recent land cover datasets have not shown substantial changes in Minnesota (see Supplementary Information, section S1.1). We refer to our modified implementation with dynamic satellite vegetation as CLM_{DS}.

The hydrological component of the model includes an upper soil column that is discretized into 10 computational layers, extending from the ground surface to about 3.8 m depth, with the thicknesses of each layer increasing exponentially with depth. Boundary conditions for the upper soil column include simulations of surface runoff, canopy and ground evaporation, infiltration, flux to and from a deeper unconfined aquifer layer when the water table is positioned below the upper soil column, and subsurface drainage when the water table is positioned within the upper soil column.

For calculating the dynamic soil water content of each computational layer within the upper soil column, CLM uses a modified form of the 1D vertical Richards equation (Zeng and Decker, 2009):

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[k\left(\frac{\partial(\psi-\psi_E)}{\partial z}\right)\right] - Q, \tag{2}$$

where θ is the volumetric soil water content, ψ is the soil matric potential, ψ_E is the equilibrium (hydrostatic) soil matric potential, Q is the soil moisture sink term (includes root uptake of water for transpiration), and k is the unsaturated hydraulic conductivity. The hydraulic parameters (moisture retention curve exponent B, saturated hydraulic conductivity k_{sat}, saturated volumetric soil moisture content θ_{sat}, and saturated matric potential ψ_{sat}) used for calculating $\psi = \psi_{sat}\left(\frac{\theta}{\theta_{sat}}\right)^{-B}$ and $k = k_{sat}\left(\frac{\theta}{\theta_{sat}}\right)^{2B+3}$ are determined by CLM based on soil texture (percent sand and clay) and organic matter content in each soil layer (Clapp and Hornberger, 1978; Cosby et al., 1984; Lawrence and Slater, 2007). Details about the pedotransfer function used in the model can be found in Oleson et al. (2013).

Determination of the water table depth and recharge is based on Niu et al. (2007). If the water table is within the upper soil column (within 3.8 m depth), recharge to the water table is calculated by applying Darcy's law across the water table:

$$q_{recharge} = -k_{aq}\frac{\psi_\nabla - \psi_{jwt}}{z_\nabla - z_{jwt}}, \tag{3}$$

where $\psi_\nabla = 0$ is the matric potential at the water table at depth z_∇, subscript *jwt* denotes the index of the layer directly above the water table, and k_{aq} is the hydraulic conductivity of the layer containing the water table ($jwt + 1$).

For the case when the water table is below the upper soil column (below 3.8 m), recharge is set to the potential recharge, defined as the net flux of water across the lower boundary of the deepest explicit soil layer at 3.8 m depth. Because most of the major water table aquifers in Minnesota are close to the surface (within 3 m; Adams, 2016), the approximation of potential recharge as actual recharge is needed for only certain parts of the state. CLM computes potential recharge by calculating the change in water content of an additional soil layer below the upper soil column:

$$q_{recharge} = \frac{\Delta\theta_{liq,N+1}\Delta z_{N+1}}{\Delta t}, \tag{4}$$

where $\Delta\theta_{liq,N+1} = \theta^{t+1}_{N+1} - \theta^{t}_{N+1}$ is the change in liquid water content in an additional lower layer (layer $N + 1$, from the bottom of the upper soil column to the water table) over a model time step (t to $t + 1$), and $\Delta z_{N+1} = z_\nabla - z_N$ is the difference between the water table depth z and the bottom of the upper soil column $z_N = 3.8m$. Liquid water content of the lowest $N+1^{th}$ layer is calculated using the modified form of Richards equation as noted above in Eq. 2, assuming no vertical flux across the bottom of the layer.

The change in water table depth is calculated based on recharge and drainage:

$$\Delta z_\nabla = \frac{q_{recharge} - q_{drai}}{S_y}\Delta t, \tag{5}$$

where Δt is the model time step, S_y is specific yield calculated as $S_y = \theta_{sat}\left[1 - \left(1 + \frac{z_\nabla}{\psi_{sat}}\right)^{-1/B}\right]$, and q_{drai} is the subsurface drainage defined as the flux of water out of the vertical soil column via lateral flux at the water table. Subsurface drainage is represented through a parameterization based on topographic statistics:

$$q_{drai} = q_{drai,max}\,exp\left(-f_{drai}z_\nabla\right), \tag{6}$$

where $f_{drai} = 2.5m^{-1}$ is a decay factor parameter, and $q_{drai,max} = 10sin(\beta)\ kg\ m^{-2}s^{-1}$ is the maximum drainage when the water table is at the ground surface, with β as the mean grid cell topographic slope (in radians). As noted previously, we consider simulations in southeast Minnesota to have higher uncertainty, largely because this parameterization is likely insufficient for representing complex flows in karst regions.

The model was run for two scenarios at a spatial resolution of 25 km (chosen for computational purposes) using a 30-min time-step over the 16-year period of 2000 to 2015 with a monthly model output frequency. In the first scenario, CLM_{CS}, the default climatological monthly LAI inputs were used. In the second scenario, CLM_{DS}, we replaced the default climatological LAI with annually dynamic LAI from the Global Land Surface Satellite (GLASS) (Xiao et al., 2014). GLASS was derived by combining MODIS and AVHRR satellite data on vegetation. It has a temporal resolution of 8-days and is available globally starting from 1982 at 1 km spatial resolution. This dataset has been validated in multiple studies and found to be in good agreement with observations (Fang et al., 2019). The LAI data was inputted to the model at its original temporal resolution of 8-day. For meteorological data inputs, we used Daymet (Thornton et al., 2014) for precipitation, temperature, and humidity, as well as NCEP NARR (Meixner et al., 2006) for wind speed and

atmospheric pressure. Soil texture inputs were obtained from POLARIS (Chaney et al., 2016). Compared to the default CLM global soil dataset, POLARIS is a much newer product that was derived from a probabilistic remapping of SSURGO (Soil Survey Staff, 2020), an observational database that specifically focuses on North America. All input datasets were regridded to the 25x25 km computational grid (details in Supplementary Information, Section S1.2). Soil hydraulic parameters derived by CLM pedotransfer functions from the soil texture inputs were further calibrated using water table depth observations from the Minnesota Department of Natural Resources's (MNDNR) Cooperative Groundwater Management (CGM) database (www.dnr.state.mn.us/waters/cgm), as described in the following section.

2.3. Model calibration with ensemble Kalman filter

As previously mentioned, simulations of vegetation-recharge coupling in CLM can be highly uncertain due to poor constraints on soil properties over the entire unsaturated zone. Because CLM is a complex nonlinear, spatially distributed model that uses a large number of parameters to simulate ecohydrological processes, model calibration is critical but also challenging. To address this, we used the Ensemble Kalman Filter (EnKF) (Evensen, 1994) to calibrate soil parameters by assimilating observations of dynamic water table depths to reduce uncertainty in recharge simulations. We focused on subsurface parameter estimation because surface model inputs of climate and vegetation variables were considered to be relatively well-known based on direct observations. EnKF was chosen for its flexible application with any model and observation type, as well as its efficient handling of high-dimensional models like CLM. Water table depth time series were selected as the calibration data because changes in water table depth capture water movement over the entire overlying soil column. EnKF combines spatiotemporally resolved details from CLM with depth-integrated information from water table depth measurements in order to generate soil property estimates throughout the soil column.

We used water table data from 320 surficial wells within the CGM database (locations shown in Fig. 2b, water level depths shown in Fig. S1 in Supplementary Information) that had repeat measurements ranging from monthly to annual frequency over the 2000–2015 simulation period. The CGM database processing procedure includes periodic checks of automated water level measurements against manual measurements for accuracy (Pearson et al., 2012). We further removed data points that showed unusually large one-time jumps in water level (more than 3.5 m within a month). Although water levels could be influenced by groundwater extractions, we did not attempt to filter for affected wells. Global studies (e.g., Gleeson et al., 2012) do not show Minnesota aquifers to be experiencing any major regional groundwater overdraws, and collecting local pumping information on a statewide scale was beyond the scope of this study.

2.3.1. Ensemble Kalman filter

EnKF is a data assimilation method that merges information from imperfect models and uncertain observations in an optimal way to reduce and quantify uncertainty (Liu and Gupta, 2007). The Kalman filter and its variants are widely used in the Earth Sciences (Carrassi et al., 2018; Liu and Gupta, 2007; Reichle, 2008; Sun et al., 2016).

A sequence of "assimilation cycles" are carried out in EnKF, each of which consists of two steps: forecast and update. In the forecast step, an N-member ensemble of state variables are evolved forward in time through the model (i.e., Monte Carlo runs), with each member using a different perturbed parameter set and/or including added state variable noise to represent model structural uncertainty. The forecast step at timestep t can be represented as:

$$x_t^f = g\left(x_{t-1}^u, p^u{}_{t-1}, b_t\right) + \omega_t, \tag{7}$$

where $x \in R^{n\times N}$ is the ensemble of n-dimensional state variables, superscripts f and u respectively denote the forecast and updated variables, $g(.)$ is the non-linear model applied to each ensemble member, $p \in R^{p\times N}$ is the ensemble of p-dimensional parameter values, b is the forcing at the boundaries, and ω includes any additional model error terms.

In the update step, when parameter estimates are desired, an augmented state matrix $\xi_t \in R^{(n+p)\times N}$ is created in which $p^u{}_{t-1}$ is appended to x_t. The ensemble members are adjusted based on observations using the following equation:

$$\xi_t^u = \xi_t^f + K\left(Y - H\xi_t^f\right), \tag{8}$$

where the matrix $Y \in R^{m\times N}$ contains a different perturbed vector of observations $y \in R^m$ in each column (with $y_j = y + \gamma_j, j = 1,\ldots,N$, and γ_j representing a random draw for a 0-mean observation error), K is the Kalman gain, and H is the measurement operator that relates the model state to the observations. The Kalman gain is computed as follows:

$$K = P_e^f H^T\left(HP_e^f H^T + R_e\right)^{-1}, \tag{9}$$

where the model error covariance is calculated as $P_e^f = \frac{(\xi_t^f - \overline{\xi}_t^f)^T}{N-1}$ with $\overline{\xi}_t^f \in R^{(n+p)\times N}$ containing replicates of the ensemble mean vector in each column, and the observation error covariance matrix is calculated as

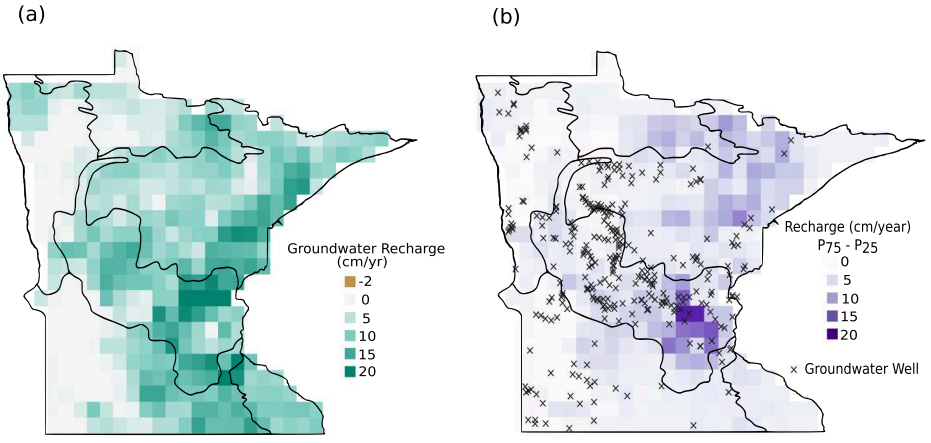

Fig. 2. (a) Estimated (ensemble median of CLM_{DS}) 16-year average (2000–2015) groundwater recharge in Minnesota. (b) Difference between 75th and 25th percentile of the ensemble recharge estimate. Darker areas indicate more uncertainty in the model outputs. × indicate locations of water table depth observations (from 320 shallow groundwater wells) used for model calibration with EnKF.

Table 2
State variables and parameters.

Variables		Description (Units)	Prior[min, max]
Soil parameters	% Sand	Fraction of sand	POLARIS ± 40%
	% Clay	Fraction of clay	POLARIS ± 40%
	B	Slope of soil water retention curve (unitless)	[2.8, 15]
	k_{sat}	Saturated hydraulic conductivity (mm sec^{-1})	[5×10^{-4}, 3×10^{-1}]
	ψ_{sat}	Saturated soil matric potential (mm)	[10, 760]
	θ_{sat}	Saturated soil water content (porosity) (mm^3/mm^3)	[0.3, 0.93]
Subsurface drainage parameters	f_{drai}	Decay factor for the subsurface drainage calculation (Eq. 6) (m^{-1})	[0.1, 15]
	$q_{\text{drai,max}}$	Maximum subsurface drainage (kg m^2 sec^{-1})	[10^{-6}, 10^{-1}]
Surface runoff parameters	f_{over}	Decay factor for the surface runoff calculation* (m^{-1})	[0.1, 5]
	f_{max}	Maximum value of saturated area of gridcell*	[0.1, 0.7]
State variable	ZWT	Water Table Depth (m)	NA

*$Surface\ Runoff = f_{max} exp(-0.5 f_{over} z_{\nabla}) q_{liq}$ where f_{max} is the potential or maximum value of fractional saturated area, which is determined by the topographic characteristics and soil moisture state of a gridcell, f_{over} is a decay factor (m^{-1}), z_{∇} is the water table depth (m), and q_{liq} is the sum of liquid precipitation reaching the ground and meltwater from snow (kg m^{-2} sec^{-1}).

$R_e = \frac{\beta\beta^T}{N-1}$ with $\beta = (\gamma_1, \gamma_2, \ldots, \gamma_N) \in R^{m \times N}$. The updated model error covariance $P_e^u = \frac{(\xi_t^u - \bar{\xi}_t^u)(\xi_t^u - \bar{\xi}_t^u)^T}{N-1}$ quantifies the uncertainty in the estimate at the end of the update step. The updated states and parameters at time t are then used for the forecast step in the subsequent assimilation cycle for time $t + 1$.

2.3.2. Implementation of EnKF with CLM

Sensitivity tests were carried out to identify the parameters most strongly controlling simulations of water table depth and groundwater recharge. Table 2 lists these parameters as well as the state variables that were estimated with EnKF using water table depth observations. EnKF was initialized at the beginning of the simulation period using random parameter draws from prior uniform probability distributions with uncertainty ranges shown in Table 2. The range for subsurface drainage and surface runoff parameters were specified based on past studies (Ray et al., 2015; Huang et al., 2016) and test simulations. Uncertainty in POLARIS dataset was accounted for by applying depth correlated soil texture perturbations similar to Ng et al. (2009) for each soil layer. The prior soil parameter range was then derived using the pedotransfer function of CLM (value range listed in Table 2) based on the perturbed soil texture inputs. Perturbation range of ± 40% for soil texture (Table 2) was chosen based on preliminary tests that showed it to be appropriate for generating ensemble simulations that cover observed water table depths.

For our implementation with CLM, we applied several modifications to the standard EnKF algorithm to improve its performance for simultaneous state and parameter estimation, which can be prone to issues due to non-linearities (Ng et al., 2014). As detailed fully in Supplementary Information (Section S1.3), these include parameter perturbations (Moradkhani, 2005), variance inflation (Anderson and Anderson, 1999), and update iterations (Liu and Ng, 2019).

We set the observation error for water table depth to 0.45 m to represent not only measurement errors but also scale representation errors between a single well and the model gridcell. Preliminary tests showed that an ensemble size of 100 provided an appropriate balance between performance and computation cost. The final set of ensemble parameter estimates from EnKF were used to execute two separate sets of model runs: CLM_{CS} using the climatological vegetation inputs, and CLM_{DS} using satellite-based dynamic vegetation as the input. The ensemble-median of CLM_{DS} simulations are taken as the final calibrated estimate, with the full ensemble representing estimation uncertainty, while CLM_{CS} simulations are used for comparison to evaluate the effect of dynamic vegetation.

3. Results and discussion

3.1. Model Performance Evaluation

To evaluate the performance of the model calibration to water table depths, we compared the simulated ensemble median water table depth and observed water table depth using the root mean square residual (RMSR) over the different observation times for 2000–2015. The spatially averaged mean RMSR improved by 51% to 1.27 m in the posterior run (after calibration with EnKF) compared to the prior (before calibration with EnKF), indicating a statewide improvement of CLM outputs (Fig. S3 in the Supplementary Information). In a 5-year model validation run over 2016–2020 using the calibrated parameters, the RMSR for water table depth remained low at 1.36 m (Fig. S4 in Supplementary Information), providing further confidence in our calibration results.

Using the calibrated model over the study period of 2000–2015, simulations of recharge show a notable difference compared to the recharge results before calibration (Fig. S5 in Supplementary Information), demonstrating the effect of the constraint on water table observations. The calibrated ensemble median recharge estimates (with dynamic vegetation) across Minnesota at 25 km resolution averaged over the 2000–2015 simulation period ranged from less than 0.1 cm year^{-1} in the western part of the state to 26.4 cm year^{-1} in the east, with a statewide average of 7.4 cm year^{-1} (Fig. 2a). This increasing west to east spatial trend agrees with past regional recharge studies (Delin et al., 2007; Smith and Westenbroek, 2015) and largely reflects the underlying precipitation pattern (Fig. 1a). The highest simulated recharge occurred in the central part of the state, which is known to contain highly conductive, glacially deposited sand plains in addition to moderately high precipitation (Smith and Westenbroek, 2015; Mesinger et al., 1993). We also found that recharge varies seasonally (Fig.S2 in Supplementary Information), with 82% of the annual statewide recharge occurring during spring (April - June), and a smaller second period of recharge occurring during fall (October and November).

Recharge estimate uncertainty was quantified using the difference between the 75th and 25th percentile of the ensemble recharge simulations in Fig. 2b. The 320 groundwater wells used for calibration data were unevenly distributed throughout the state with almost all in central, northwestern, and southwestern Minnesota, allowing for lower uncertainty in many of those areas (Fig. 2b). Higher uncertainty in the recharge estimate in northeastern Minnesota corresponds with where there is almost no water table depth data. The high uncertainty in central Minnesota is around the Minneapolis-St.Paul Twin Cities Metropolitan Area. It is likely that uncertainties persist there despite the availability of water table depth observations, because CLM inadequately captures the effects of urbanization and management on groundwater, which is the source of more than 70% of the municipal

water used in the Twin Cities metro area (Twin Cities Metropolitan Council, 2007).

Our simulated statewide recharge estimates (the ensemble median) were generally lower (~ 35% statewide) than previous water balance-based recharge estimates in Minnesota (Smith and Westenbroek, 2015). Because direct observations of recharge are lacking for validation, we compared our estimates to available observation-based data for other major hydrological fluxes, specifically ET and runoff. Evaluating ET is particularly important for our study given the focus on vegetation. Our statewide-average simulated ET closely matched (within 5%) the ET product from MODIS (MOD16 (Mu et al., 2007; Mu et al., 2011)) and fell between two other ET products that use satellite and flux tower measurements (SSEBop (Senay et al., 2013) and GFET (Jung et al., 2009)) (Fig. S6 in Supplementary Information). Generally, locations with moderate to high ET simulations better match the observation-based ET products compared to locations with low ET simulations (Fig. S7 in Supplementary Information). We also compared ensemble median streamflow (calculated as subsurface drainage plus surface runoff in

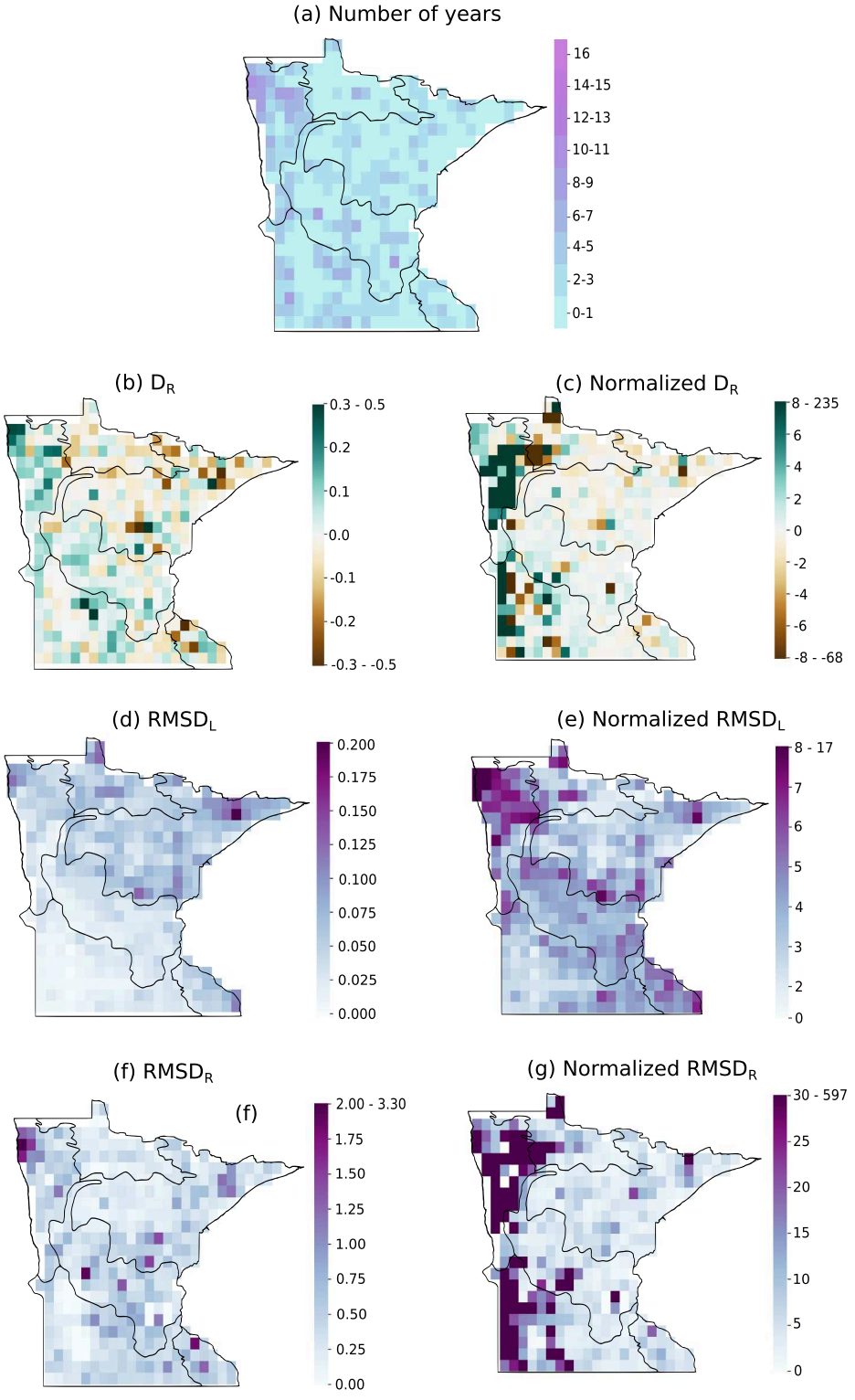

Fig. 3. (a) The number of years (out of 16) when the absolute value of the recharge difference between CLM_{DS} and CLM_{CS} was greater than 10%. (b) Average difference between CLM_{CS} and CLM_{DS} for recharge (D_R, Eq. 10) (cm/yr) (c) normalized D_R (%) (d) Root mean squared deviation between CLM_{CS} and CLM_{DS} for LAI inputs ($RMSD_L$, Eq. 12) (e) normalized $RMSD_L$ (%) (f) recharge ($RMSD_R$, Eq. 11) (cm/yr) (g) normalized $RMSD_R$ (%). The colorbar label includes the minimum and. maximum values.

CLM) simulations with measurements from 13 long-term USGS stream gauges and found that the model captured overall observed spatial trends without any statewide bias (Fig. S8 in Supplementary Information). Reasonable estimation of these major components of the water balance as well as water table depth gives confidence in the model's ability to satisfactorily simulate groundwater recharge and its links with vegetation dynamics. Although errors could persist in absolute groundwater recharge estimates from CLM, these likely become less important when we examine differences in recharge for dynamic versus climatological vegetation inputs.

3.2. Impacts of dynamic vegetation on recharge

3.2.1. Statewide results

Below, we used the ensemble medians of CLM_{DS} and CLM_{CS} simulations to evaluate the impacts of dynamic vegetation on recharge, as well as the influence of climatic drivers. The statewide recharge simulated with dynamic satellite vegetation (CLM_{DS}) was very similar to the result with climatological satellite vegetation (CLM_{CS}) (7.4 versus 7.1 cm/yr). This similarity indicates that across the state over the 16-year period, times and locations with higher than average recharge with dynamic vegetation were mostly offset by other times and locations with lower than average recharge. As such, dynamic vegetation seems not to be important to consider for climatological studies of recharge.

However, as reflected in the overall spatiotemporal correlation of -0.4 (p-value < 0.05) between LAI differences and simulated annual recharge differences for CLM_{DS} versus CLM_{CS}, higher (lower) than time-averaged LAI conditions can cause a decrease (increase) in recharge in certain years and locations. Correspondingly, Fig. 3a shows the occurrence of appreciable differences between recharge results with and without dynamic vegetation over the 16-year period. Here, appreciable differences are defined as those years when the difference in recharge is at least 10% (calculated relative to CLM_{CS}) and the average annual recharge is at least 10 mm/yr. These criteria were set to detect cases in which recharge differences may be sufficiently large as to have water resource and ecological implications. Across different grid cells, an average of 2 years out of the 16-year simulation period experienced appreciable differences in annual recharge simulations. The drier northwest showed the greatest recharge sensitivity to dynamic vegetation, with an average 12 out of 16 years showing appreciable differences.

To evaluate the size of recharge differences over the 16 years, and to compare them to the underlying dynamic vegetation conditions, we first calculated the time-averaged recharge difference (D_R) for each grid cell over the simulation period:

$$D_R = \frac{\sum_t \left(R(t)_{DS} - R(t)_{CS} \right)}{n} \tag{10}$$

where t is the time index for simulation year, $n = 16$ is the number of years simulated, and $R(t)$ is the simulated annual recharge at time t for the CLM_{DS} or CLM_{CS} scenario. Even though average LAI differences over the 16 years between the scenarios is zero (because climatological vegetation is defined as the mean of the 16-year dynamic vegetation), we found that a number of grid cells exhibit a sign bias in D_R (Fig. 3b), ranging from 0.5 cm/yr to -0.5 cm/yr. The notable non-zero D_R results in some grid cells indicate that cumulatively over the 16-year simulation period, there can be an asymmetric impact on recharge by either higher or lower than average LAI conditions, resulting in respectively lower or higher overall recharge than expected with climatological vegetation conditions. For example, in Fig. 3b, recharge in parts of northwestern and southern Minnesota on average experienced higher recharge over the 16 years (positive D_R), indicating greater influence by lower-than-average LAI. In contrast, a number of locations in central and eastern Minnesota with lower than expected recharge over the 16 years (negative D_R) show greater sensitivity to above-normal LAI. While D_R magnitudes appeared heterogeneously distributed across the state in Fig. 3b, recharge differences normalized by time-averaged local recharge in Fig. 3c were highest in the western part of the state, where recharge values are lowest (Fig. 2a). These differences ranged from a minimum of -68% to a maximum of $+235\%$ of average recharge for different grid cells, indicating that dynamic vegetation can have a substantial relative impact on recharge over a 16-year period.

While Fig. 3b shows whether there is a tendency for higher or lower recharge to accumulate over the 16 years due to dynamic vegetation, within a grid cell, annual recharge in CLM_{DS} likely falls both above and below what is expected with climatological vegetation over the 16 years. To examine the magnitude of recharge and LAI differences between the CLM_{DS} and CLM_{CS} scenarios regardless of the sign, we calculated the root mean square difference (RMSD) over the 16-year period, i.e.

$$RMSD_R = \sqrt{\frac{\sum_t \left(R(t)_{DS} - R(t)_{CS} \right)^2}{n}} \tag{11}$$

$$RMSD_L = \sqrt{\frac{\sum_t \left(L(t)_{DS} - L(t)_{CS} \right)^2}{n}} \tag{12}$$

where $L(t)$ is the input LAI at t for the CLM_{DS} or CLM_{CS} scenario, and we show non-normalized and normalized results in Figs. 3d-g. By comparing Fig. 3b, d, and f, it can be seen that locations with higher $RMSD_R$ and $RMSD_L$ often correspond to areas with greater (positive or negative) D_R, such as north-western Minnesota. However, it can also be seen that as expected (because of both higher and lower recharge over the 16 years), average recharge difference magnitudes ($RMSD_R$) are generally much greater than average difference values (D_R) over the 16-year simulation period (compare values in Figs. 3.b and f). Spatially averaged over the state, $RMSD_R$ and $RMSD_L$ normalized respectively by mean recharge and LAI were 21.1% and 4%, and for some parts of the state, they respectively reached as high as 597% and 17% (Fig. 3e and g). The higher RMSD for recharge compared to LAI demonstrates that year-to-year variability in vegetation growth may drive disproportionately high relative variability in recharge.

3.2.2. Ecoregion-specific analysis

The different major ecoregions of Minnesota (Fig. 1b) may be expected to exhibit varying sensitivity of recharge to interannual changes in vegetation growth, because of their differences in climate and plant type. Although recharge differences over the 16 years show substantial spatial variability between grid cells in Fig. 3b (spatial coefficient-of-variations of 16), some alignment with ecoregions can be discerned. For example, it was previously noted that northwest Minnesota, corresponding to the NGP ecoregion, shows greater frequency and magnitudes of recharge differences for CLM_{DS} compared to CLM_{CS} (Fig. 3a, b, f) as well as greater normalized LAI differences (Fig. 3e). Ecoregion-based trends become clearer when evaluating transient annual LAI and recharge values in Figs. 4 and 5, rather than aggregating over the 16-year simulation period (as was done for average differences and RMSD).

Here, we use D^A_R and D^A_L to denote annual level differences in recharge and LAI, respectively, between CLM_{DS} and CLM_{CS}; these are defined analogously to D_R in Eq. 10 except that they are not averaged over time. Note that because LAI inputs in CLM_{CS} are the long-term average of LAI inputs in CLM_{DS}, D^A_L represents LAI anomalies in each year (differences from the temporal average). In Fig. 4, we show D^A_L normalized by the 16-year mean LAI ($\widetilde{D}^A_L$, where the tilde indicates normalization) in the respective grid cell, because normalized D^A_L exhibits a stronger spatiotemporal correlation with D^A_R across the state compared to non-normalized D^A_R (-0.47 versus -0.4). The stronger correlation likely occurs because the normalization accounts for the

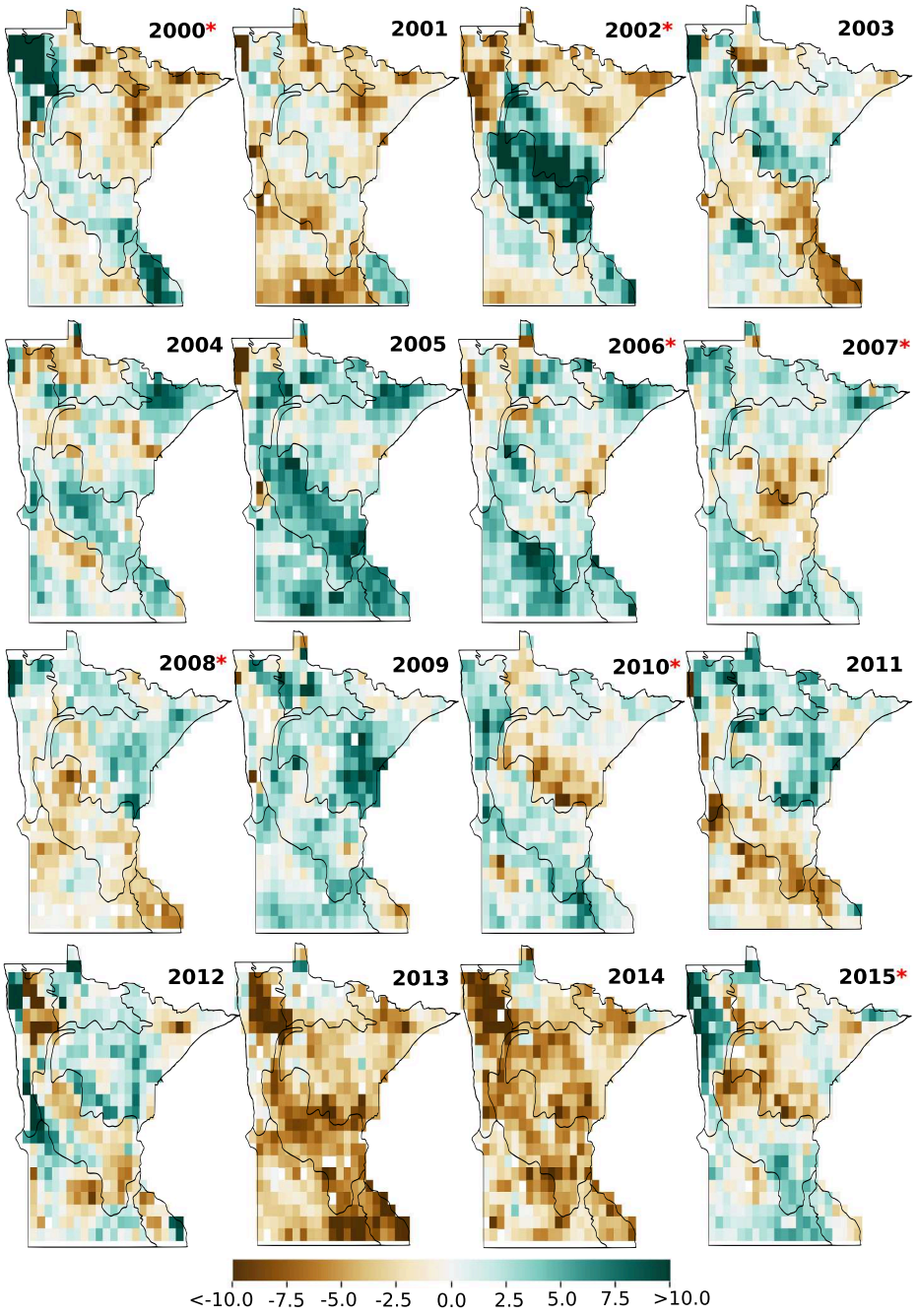

Fig. 4. Normalized difference between CLM_{CS} and CLM_{DS} for LAI inputs ($\widetilde{D}_L^A$) (%). Red asterisk (*) denotes years when ecoregion was a statistically significant predictor of $\widetilde{D}_L^A$ ($R^2 >= 0.3$; p-value < 0.05).

underlying diversity in LAI due to plant type (see Fig. S9 in Supplementary Information for PFT information). Clearest ecoregion trends can be seen in $\widetilde{D}_L^A$ in Fig. 4, for which ecoregion is a statistically significant predictor in nearly half the years using a simple regression test (p-value < 0.05). For example, Fig. 4 shows significant regional greening (positive $\widetilde{D}_L^A$) occurred in CHF in 2002, and moderate regional greening occurred in NF in 2008. These results suggest that even though some years experienced similar $\widetilde{D}_L^A$ across the state, such as 2013 and 2014, LAI variability is often influenced by ecoregion-based conditions.

The corresponding recharge differences between CLM_{DS} and CLM_{CS} in Fig. 5 show considerably more spatial heterogeneity than LAI anomalies, as seen in the variable results among neighboring grid cells with similar LAI anomalies. A regression test revealed that ecoregion was not a significant predictor for recharge differences for any of the 16 years (p-value > 0.05). However, a negative response of recharge to the LAI differences can still be gleaned through strong, statistically significant temporal correlations between $\widetilde{D}_L^A$ and D_R^A of -0.68 to -0.85 for the different ecoregions (Fig. 6a). This relationship can be seen, for example, in the uniformly low LAI years of 2013 or 2014, when recharge was generally higher than expected with climatological vegetation in all four ecoregions, as well as in 2002 and 2008, when regional greening corresponded with mostly reduced recharge (Fig. 5). As a result, some correspondence between ecoregion and recharge differences does appear based on underlying ecoregion controls on LAI anomalies.

In Fig. 7a-b, boxplots of $\widetilde{D}_L^S$ and $D^S{}_R$ summarize how much LAI and corresponding recharge differences vary across the 16 years for each

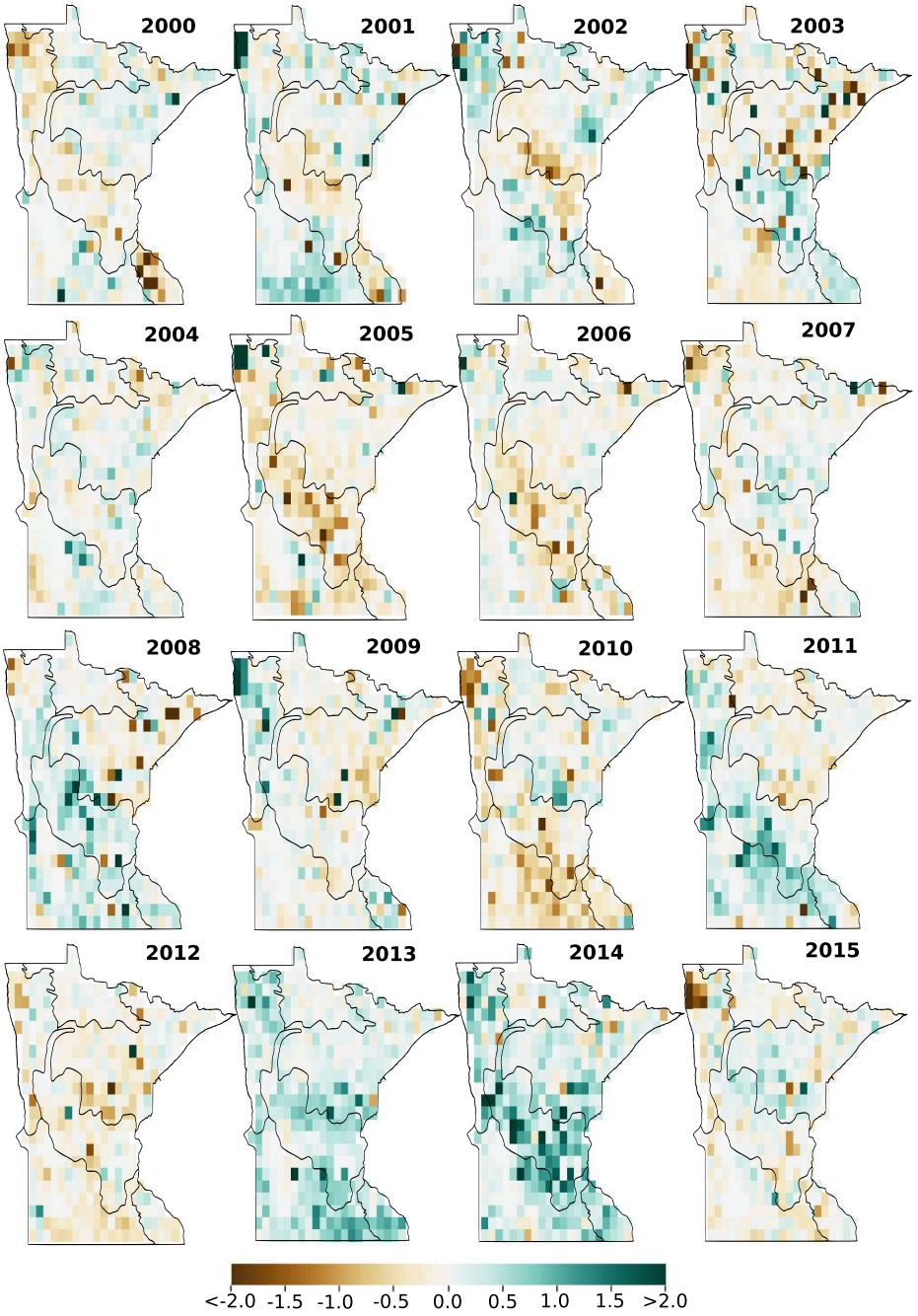

Fig. 5. Difference between CLM_{CS} and CLM_{DS} for recharge (D_R^A) (cm $year^{-1}$). Positive values indicate a higher value in CLM_{DS}.

ecoregion (superscript S here denotes seasonal average). By further breaking these results down by season, we note that LAI anomalies not only differ by ecoregion, but they also vary within a year (Fig. 7a). Winter (December-March) is omitted from our discussion, because LAI anomalies and corresponding recharge differences between CLM_{DS} and CLM_{CS} are very small during that season. It can be seen that the greatest normalized differences (positive and negative) between dynamic and climatological LAI occur in the spring (April-June) for all ecoregions except NF. The large spring-time variability in NGP, SGP, and CHF reflects the year-to-year differences in leaf-out or crop dates. In NF, where there are substantial evergreen forests (Table 1), spring-time LAI anomalies are smaller and more similar to other times of the year. Fall (October and November) shows the second largest LAI differences between CLM_{DS} and CLM_{CS}, due to year-to-year differences in senescence timing, especially in NGP, SGP, and CHF. The spring-dominated seasonal pattern in normalized LAI differences is translated to spring-dominated recharge differences in NGP, SGP, and CHF (Fig. 7b). This is also reflected in the correlations between $\widetilde{D}_L^S$ and $D^S{}_R$ parsed out by season, which show that recharge is most consistently sensitive to LAI anomalies in the spring (correlations of −0.80 to −0.92 across ecoregions, Fig. 6b). In springtime, both $D^S{}_R$ and $\widetilde{D}_L^S$ are greatest in the two Great Plains ecoregions, NGP and SGP, followed by CHF (Fig. 3b). This could reflect crop management decisions in the Great Plains (Table 1), which can differ more dramatically year-to-year than natural shifts in leaf-out, with major implications for recharge. Even though CHF and NGP have similar crop and grassland land cover, lower $\widetilde{D}_L^S$ in CHF might be because of the higher average LAI of temperate forests in CHF

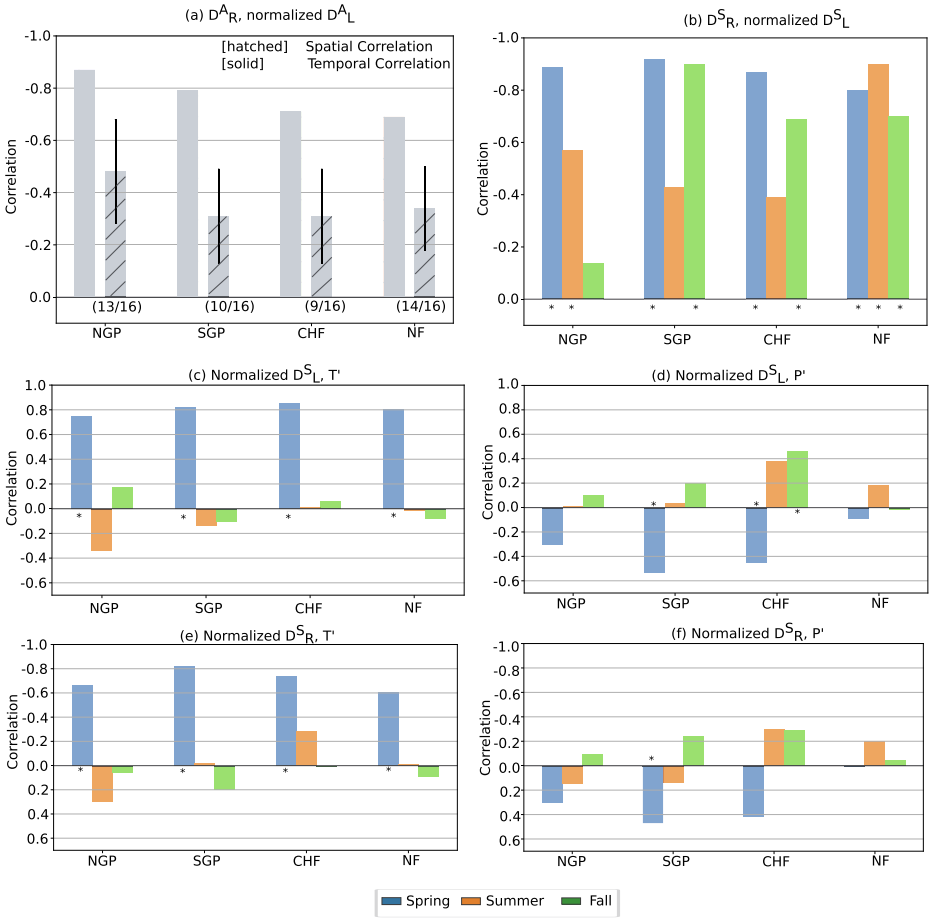

Fig. 6. (a) Solid bars show temporal correlations between ecoregion-averages of D^A_R and $\widetilde{D}^A_L$; all temporal correlations were statistically significant (p-value < 0.05). Hatched bars show 16-year means of the spatial correlations between D^A_R and $\widetilde{D}^A_L$ within each ecoregion, with error bars denoting $\pm$ 1 standard deviation over the 16 years, and with numbers below the solid bar indicating the years out of 16 when the spatial correlation was statistically significant (p-value < 0.05). (b) Seasonal temporal correlation between D^S_R and $\widetilde{D}^S_L$. Asterisk (*) denotes statistically significant correlation (p-value < 0.05). (c, d, e, f) Temporal correlation between ecoregion-averaged $\widetilde{D}^S_L$ and T′ (c), $\widetilde{D}^S_L$ and P′ (d), D^S_R and T′ (e), and D^S_R and P′ (f). Asterisk (*) denotes statistically significant correlation (p-value < 0.05). Note that the y-axis is reversed in (e) and (f) to facilitate comparisons with (c) and (d). Spring: April-June, summer: July-September, fall: October-November.

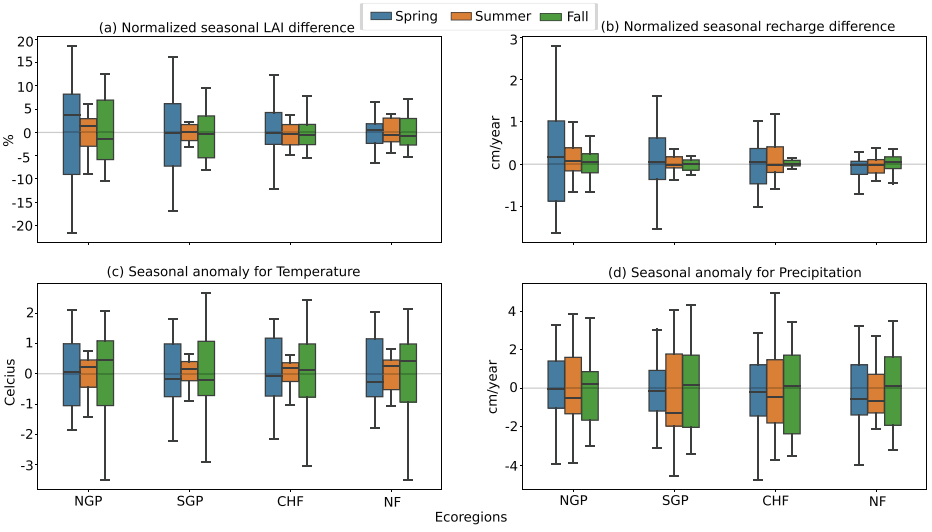

Fig. 7. Seasonal differences between CLM_{DS} and CLM_{CS} (CLM_{DS}-CLM_{CS}) in each ecoregion for (a) $\widetilde{D}^S_L$ and (b) recharge (D^S_R). Seasonal (mean over 16 years) anomalies for (c) air temperature (T′) and (d) precipitation (P′). Boxplots show the temporal spread over 16 years within each ecoregion. Spring: April-June, summer: July-September, fall: October-November. Winter is omitted because LAI and recharge differences then are inconsequential.

compared to boreal forests in NGP, which can lower the normalized LAI difference. Interestingly, moderately high $\widetilde{D}^S_L$ in the fall in NGP, SGP, and CHF is not reflected in higher recharge differences (Fig. 7a-b); this is further discussed in Section 3.3.3.

3.3. Controls on vegetation dynamics and recharge response

3.3.1. Impact of ecohydrological partioning

To evaluate the ecohydrological processes driving the recharge responses to LAI anomalies, we further examined model simulations of ET and surface runoff. Note that interception was not included in our

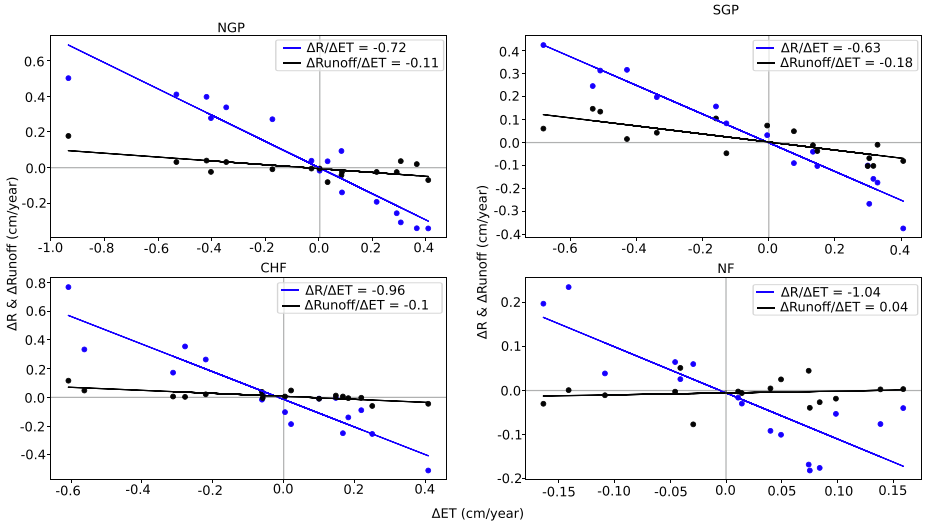

Fig. 8. Scatter plot of ecoregion-averaged ΔR (recharge) and ΔRunoff (surface runoff) versus ΔET for each year. Δ values are calculated as CLM_{DS} - CLM_{CS}. Note that ΔET is not exactly equal to ΔR + ΔRunoff because the median of the 100-member ensemble simulations is taken for each variable. Individual ensemble model simulations do satisfy water balance.

analysis because its simulated magnitude was negligible compared to ET and surface runoff. ET typically increases with higher LAI and can result in a lesser amount of water reaching the water table as recharge (Kergoat, 1998; Vivoni, 2012). We found very strong temporal correlations between annual simulations of ecoregion-averaged LAI anomalies and corresponding differences in ET (calculated analogously as D_R^A) that ranged from 0.85 to 0.91 (p-value < 0.05). As reflected in the negative temporal correlations of −0.68 to −0.85 noted above between $\widetilde{D}_L^A$ and D_R^A, an increase in ET due to greater LAI can cause a decrease in recharge, because there is less water that can pass below the root zone. However, how much a difference in annual ET (for CLM_{DS} versus CLM_{CS}) ultimately affects recharge depends on how the ET difference first influences surface runoff in the upper part of the soil column. In fact, the partitioning of ET differences to surface runoff in addition to recharge explains the weaker temporal correlation for LAI differences with recharge differences than with ET differences. Differences in net water storage change can also affect the recharge response to ET change, but storage changes were found to be small on an annual timescale.

How ET differences affect surface runoff can be complex. Lower soil moisture with greater ET can reduce saturation-excess surface runoff when conditions are generally wet, but it may increase infiltration-excess surface runoff when conditions are generally drier. We determined the average partitioning of ET differences between recharge and surface runoff differences by fitting lines through annual simulated values over the 16 years in each ecoregion (Fig. 8). We found that much of the ET differences are partitioned to recharge (63–100%, see slopes reported in figure legends), but how much differs by ecoregion (Fig. 8). While nearly all of ET differences are reflected in recharge differences in CHF and NF, the lowest fraction occurs for SGP, which has the highest partitioning of ET differences to surface runoff. The stronger influence on surface runoff in SGP, compared to other ecoregions, results from the high surface runoff to total runoff ratios values in that region (Fig. S11 in Supplementary Information), driven by higher surface runoff soil parameters (f_{max} in Fig. S13 in Supplementary Information) and less precipitation. Hydrogeologic and climatic factors are further considered in the next sections. Our findings on ET partitioning show that even though statistically, ecoregion does not appear to be a very strong predictor for recharge differences with vegetation dynamics (noted above in Fig. 5), distinct hydrological processes based on ecoregion do in fact affect the response of recharge to LAI and ET changes.

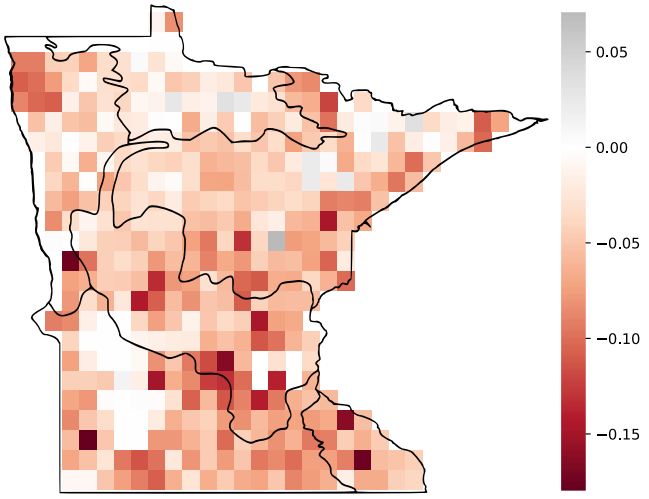

Fig. 9. Slope of regression over time (m_{RL}) to predict the recharge difference between CLM_{CS} and CLM_{DS} (D_R^A) based on the $\widetilde{D}_L^A$. Negative values indicate that higher than normal LAI corresponds with a recharge estimate that is lower than that predicted with climatological LAI inputs.

3.3.2. *Impact of hydrogeologic factors*

Compared to temporal correlations between ecoregion-averaged $\widetilde{D}_L^A$ and D_R^A (-0.68 to -0.85), spatial correlations within each ecoregion between $\widetilde{D}_L^A$ and D_R^A are much weaker (-0.31 to −0.48) (Fig. 6a). This can also be seen in the high grid cell-to-grid cell variability in D_R^A compared to $\widetilde{D}_L^A$ (Figs. 4 and 5). The stronger spatial variability in D_R^A suggests that although LAI anomalies overall drive recharge differences between CLM_{DS} and CLM_{CS} through changes in ET, the specific response is influenced by local heterogeneities such as in hydrogeologic properties, as we already started to see in the previous section. To analyze spatial variability in how recharge responds to dynamic vegetation, we conducted a regression over time to predict D_R^A based on $\widetilde{D}_L^A$ for each gridcell and then evaluated the slope of the regression line (denoted as $m_{R,L}$). The value of the slope provides a measure of recharge sensitivity to vegetation dynamics by quantifying the change in D_R for a unit change in normalized D_L (negative slope indicates an inverse relationship).

Table 3
Regression analysis results for predicting $m_{R,L}$ using normalized predictor variables. Predictor variables are explained above in Table 2 and Section 2.2. Greater coefficient value (positive or negative) indicates a stronger predictor. Higher t-value (positive or negative) indicates greater confidence in the variable as the predictor.

	R^2	Predictor (statistically significant; p-value < 0.05)	Coefficient (for standardized predictor variables)	t-statistic
NGP	0.5	B	−0.45	−2.5
		k_{sat}	−0.61	−2.2
		f_{over}	−0.42	−2.6
SGP	0.4	f_{max}	0.5	3.6
		f_{drai}	−0.4	−3
CHF	0.4	k_{sat}	−0.42	−3.7
		f_{max}	0.39	0.3
		f_{drai}	−0.2	−2
NF	0.1	None		

Table 4
List of abbreviations and terms.

Abbreviations & terms	**Explanation**
CGM	Cooperative Groundwater Management
CHF	Central Hardwood Forests
CLM	Community Land Model
CLM_{CS}	CLM Climatology Satellite Vegetation
CLM_{DS}	CLM Dynamic Vegetation
D^A_L	Annual LAI difference
$\widetilde{D}^A_L$	Normalized D^A_L
D^A_R	Annual Recharge difference
D_R	Time averaged recharge difference
D^S_L	Seasonal LAI difference
$\widetilde{D}^S_L$	Normalized D^S_L
D^S_R	Seasonal recharge difference
EnKF	Ensemble Kalman Filter
ET	Evapotranspiration
LAI	Leaf Area Index
LU/LC	Land use/Land cover
MNDNR	Minnesota Department of Natural Resources
$m_{R,L}$	Slope of regression line for D^A_R and $\widetilde{D}^A_L$
NF	Northern Forests
NGP	Northern Great Plains
P′	Precipitation anomalies from 16-year mean
PFT	Plant functional types
RMSD	Root mean square difference
SGP	Southern Great Plains
T′	Temperature anomalies from 16-year mean

By examining the spatial distribution of $m_{R,L}$ in Fig. 9, some regional patterns emerge. For example, the northwestern corner of NGP shows a cluster of high sensitivity (strongly negative $m_{R,L}$) where normalized LAI anomalies are also generally greatest (Fig. 3e), while SHP includes areas of the lowest sensitivity where absolute recharge amounts are also very low (Fig. 2a). However, as can be expected from lower spatial correlations between $\widetilde{D}^A_L$ and D^A_R compared to temporal, the regression slope exhibits considerable grid cell-to-grid cell variability through all ecoregions.

To assess the spatial influence of hydrogeologic parameters on the sensitivity of recharge to changing vegetation, we carried out a multiple linear regression analysis on the regression slope ($m_{R,L}$). Specifically, we used the calibrated soil hydraulic parameters, subsurface drainage parameters, and surface runoff parameters listed in Table 2 (and shown in Fig. S13 in Supplementary Information) as predictors of $m_{R,L}$ for each ecoregion. k_{sat} and θ_{sat} were the only highly correlated pair of parameters (correlation > 0.8; p-value < 0.05), and so we dropped θ_{sat} from our regression model to remove effects of multicollinearity but retained all other parameters. To allow direct comparison of regression coefficients in Table 3, we also standardized the predictor variables (subtracted the mean and divided by the standard deviation) before conducting the regression. In Table 3, we see a good overall multiple linear regression model fit with R^2 values ranging from 0.4 to 0.5 for NGP, SGP and CHF, indicating that the hydrogeologic parameters are able to explain significant variability in $m_{R,L}$ for these ecoregions.

Interestingly, the statistically significant predictors of $m_{R,L}$ were found to differ in the various ecoregions, indicating that distinct hydrogeologic processes affect the sensitivity of recharge to LAI anomalies based on ecoregion. NGP and CHF had k_{sat} as the most important predictor for the regression slope, while it was the surface runoff parameter f_{max} for SGP (based on coefficient magnitudes in Table 3). This result is consistent with the above finding that SGP has the greatest fraction of ET changes partitioned to surface runoff, which subsequently affects how much recharge is impacted. f_{max}, which positively controls surface runoff, shows the highest calibrated values in SGP (Fig. S13 in Supplementary Information). Importantly, these calibrated f_{max} values are much higher than the pre-calibration, topographically determined f_{max} values. This suggests that the calibrated values are likely reflecting the effect of the extensive tile drainage system present throughout southwestern Minnesota, which were installed to manage soil moisture by moving water from the shallow soil column to surface water features (Smith et al., 2018). Although the use of tile drains is also common in the northern Great Plains, their impact may be more subtle there because of lower precipitation and already high runoff due to clayey soils. In ecoregions where more of the ET differences are partitioned to recharge, such as CHF and NGP, k_{sat} is important for predicting $m_{R,L}$, because it directly controls flux through the soil column. CHF also includes relatively high values of calibrated f_{max} (Fig. S13 in Supplementary Information), which appears as another major predictor of $m_{R,L}$ in that ecoregion (Table 3). Note that for all ecoregions with good R^2 values in Table 3, there was more than one statistically significant soil parameter predictor, which explains greater spatial variability in $m_{R,L}$, and consequently D^A_R, than in any single soil parameter.

The poor predictive model for $m_{R,L}$ in NF could be due to variable conditions in this large ecoregion. For example, compared to other ecoregions, NF exhibits much greater variety in primary and secondary PFT assignments (Fig. S9 in Supplementary Information). PFT variables were not included in the multiple linear regression analysis but could drive some of the spatial variability in recharge responses to LAI anomalies in NF, especially given differences in phenology and rooting depths among vegetation types.

With $\widetilde{D}^A_L$ corresponding strongly with ecoregions (Fig. 4), it is likely that any correspondence between ecoregion and simulated recharge response to dynamic vegetation is mostly driven by underlying patterns in LAI anomalies. However, as the above multiple regression analysis shows, the recharge response is also influenced by spatially variable hydrogeologic parameters, and the importance of these parameters varies depending on the ecoregion.

3.3.3. Impact of climatic factors

To assess the drivers of the year-to-year vegetation variations that impact recharge, we examined their relationships with the major climatic factors of air temperature and precipitation. Across the four ecoregions, temperature anomalies from the 16-year mean (T′) are greatest in the fall and spring (Fig. 7c), while precipitation anomalies from the 16-year mean (P′) show more similarity across seasons and ecoregions (Fig. 7d). High spring-time temporal correlations between temperature and LAI in Fig. 6c suggest that in spring, notable temperature anomalies strongly drive LAI anomalies across all ecoregions, showing the well-known effect of warm spring temperatures triggering leaf-out and influencing cropping schedules (Jeong et al., 2011; Wang et al., 2011; Los, 2012). This interaction results in the large spring-time normalized LAI anomalies found in NGP, SGP, and CHF, though its effect

is tempered in NF (Fig. 7a), where the magnitude of LAI anomalies are likely muted by the evergreen forest fraction, as previously noted. Although spring-time T′ ranges are similar across ecoregions, corresponding spring-time $\widetilde{D}_L^S$ are greater in NGP and SGP, in part because of the effect of smaller nominal LAI values there (Fig. S12 in Supplementary Information) on the normalized anomalies. Interestingly, fall-time correlations between $\widetilde{D}_L^S$ and T′ are very weak (Fig. 6c), indicating that although cold temperatures in addition to photoperiod induce leaf-fall, the sizable temperature anomalies in fall are not the main control on fall LAI anomalies. Weak summer-time correlations between $\widetilde{D}_L^S$ and T′ can be expected, given that between leaf-out and leaf-fall/harvest, LAI is relatively constant between years for mature forests/grasslands and established croplands.

Fig. 6d shows that in contrast to temperature, precipitation anomalies have correlations with LAI anomalies that are not statistically significant across all ecoregions. This was not surprising in the wetter NF region, where near-zero correlations in all seasons reflect consistently well-watered conditions in the eastern part of the state (Fig. 1). In the more arid NGP and SGP regions, modest negative spring-time correlations between precipitation and LAI anomalies could indicate less plant water stress during cooler times. However, lack of statistical significance in these correlations as well as very small positive correlations between summer precipitation anomalies and LAI anomalies makes this difficult to conclude. Statistical significance for P′ and $\widetilde{D}_L^S$ correlations was found for SGP and CHF spring, when the correlation with temperature was still stronger, and for CHF fall, when the effect of dynamic vegetation on recharge is small (Fig. 7b). The cross-ecoregion dominance of temperature rather than precipitation correlation with LAI anomalies shows that despite the precipitation gradient across the state, the cold continental climate makes spring temperature the primary factor controlling year-to-year LAI variability. Our finding of temperature being the primary driving factor of vegetation changes is consistent with several studies that have investigated the response of vegetation to changing climatic factors (Wu et al., 2015; Quetin and Swann, 2018; Braswell, 1997; Piao et al., 2006). For example, Wu et al. (2015) found vegetation growth in the middle to high latitudes (30°N-90°N) was most strongly correlated with temperature changes among climatic variables that included precipitation and solar radiation. Similarly, Jeong et al. (2011) showed that climate warming resulted in an extended vegetation growing season in temperate forests in the Northern Hemisphere, and an increase in vegetation greenness during the growing season.

Like $\widetilde{D}_L^S$, $D^S{}_R$ shows the strongest relationship with temperature anomalies in the spring, compared to precipitation and other seasons (Fig. 6e-f), which is consistent with large (negative) correlations between $\widetilde{D}_L^S$ and $D^S{}_R$, especially in the spring (Fig. 6b). This can be explained by warm spring temperatures supporting higher LAI anomalies, which in turn decreases recharge through increased ET (Vivoni, 2012). Temperature anomalies outside of spring, as well as precipitation anomalies in all seasons, often also show similar relationships with $D^S{}_R$ as with $\widetilde{D}_L^S$ (compare Figs. 6c and d with e and f, respectively). Notably, the relationships are always considerably weaker than the spring-time relationship with temperature. However, there are occasionally distinctions in the climate anomaly correlations with $D^S{}_R$ versus $\widetilde{D}_L^S$ during summer and fall in NGP, SGP, and CHF, as reflected in the weaker $D^S{}_R$ and $\widetilde{D}_L^S$ correlations for these instances (Fig. 6b). These weaker summer and fall correlations between $D^S{}_R$ and $\widetilde{D}_L^S$ could be appearing because of water stress in these seasons, outside of the humid NF ecoregion. Water stress can dampen the influence of LAI on ET (and consequently on recharge), such that greater LAI may not support proportionately higher ET due to weather limitations. Water stress in the summer and fall outside of NF can be seen in the low (sometimes negligible or negative) recharge values in these ecoregions (see Fig. S2 in Supplementary Information). For this study, summer and fall processes are of less concern for vegetation-driven recharge variability, given that the magnitudes of $\widetilde{D}_L^S$ and $D^S{}_R$ are generally smaller in those seasons (Fig. 7a). Overall, although non-spring seasons can exhibit appreciable $D^S{}_R$ values (Fig. 7), apart from in NF, $D^S{}_R$ magnitudes are greatest in the spring, when temperature is the clear driving factor through its control on LAI anomalies (Fig. 6c).

Our finding about the importance of temperature, compared to precipitation, for recharge variability may be surprising, because precipitation is of course the primary driver of absolute recharge amounts. For example, the spatial distribution of 16-year average recharge across the state is strongly correlated with mean annual precipitation (0.86, p-value < 0.05), while it shows little correspondence with mean temperature (0.07; p-value > 0.05). It is important to note that climate inputs were identical in our two model scenarios, which only differed in their LAI inputs. Our modeling study thus does not attempt to demonstrate direct effects of climate variability on recharge. Instead, it brings to light the role of vegetation in mediating the influence of climate anomalies on recharge variability. While precipitation serves as the first-order driver of recharge amounts, we showed that year-to-year changes in vegetation variables such as LAI, which in Minnesota are largely driven by spring temperatures, further impact recharge. This impact on recharge occurs on the magnitude of about 0.4 cm/yr or up to 40% in certain ecoregions over the 16-year study period (Fig. S10 in Supplementary Information).

4. Summary and conclusion

Our study investigated the impact of dynamic vegetation on groundwater recharge by comparing CLM model simulations for two different vegetation scenarios: CLM_{DS}, in which satellite-based, year-to-year varying LAI conditions were implemented as model inputs, and CLM_{CS} (the standard approach) in which climatological (16-year average) monthly LAI conditions were implemented. The study was carried out for 2000–2015 at a resolution of 25 km for the state of Minnesota. Focusing on the surface and subsurface hydrological properties, we used EnKF to calibrate runoff, subsurface drainage, and soil hydraulic parameters by assimilating statewide observations of dynamic water table depth in order to reduce uncertainty in the recharge simulations.

We found that 16-year statewide average recharge was similar when considering dynamic versus climatological vegetation, indicating that year-to-year varying vegetation was not important for climatological estimates of recharge. However, we did find notable responses in recharge to dynamic vegetation on an annual scale. Statewide, the average magnitude of annual recharge differences with and without dynamic vegetation ($RMSD_R$) was 21.1%. This recharge difference corresponded to only an average 4% average magnitude difference between climatological and dynamic LAI ($RMSD_L$), indicating a disproportionately high impact on recharge by year-to-year variability in vegetation.

As reflected in a strong negative temporal correlation across the state between normalized LAI anomalies and simulated recharge differences for the two model scenarios, higher-than-normal LAI reduces groundwater recharge due to higher ET. How much recharge is affected by LAI anomalies through ET depends on both larger scale ecoregion conditions and local scale hydrogeological conditions. In many of the years, LAI anomalies align with the major ecoregions of Minnesota, and ecoregions with the greatest normalized springtime LAI anomalies also exhibit the greatest differences in springtime recharge, which constitutes the bulk of annual recharge. The drier ecoregions in western Minnesota (NGP, SGP) undergo the greatest normalized LAI anomalies and total recharge differences with dynamic vegetation, while the wet eastern part of the state (NF) shows much stabler interannual LAI and corresponding recharge differences. Ecoregion also appears to determine how much LAI-driven ET differences are partitioned to recharge versus runoff in the water balance. In most ecoregions, nearly all of ET differences is reflected in

recharge differences, but in southwestern Minnesota (SGP), a substantial amount of excess water that results from reduced ET is likely removed by tile drains before reaching the water table.

High within-ecoregion spatial variability in recharge differences (for the two scenarios) reveal that local hydrogeological conditions also control the response of recharge to LAI anomalies. Using multivariable linear regression tests, we found that hydrogeologic parameters were able to explain significant variability (40% to 50%) of the sensitivity of recharge to dynamic vegetation within three out of four ecoregions. NF had more varied vegetation types (PFTs) than other ecoregions, likely contributing to the poor predictability there of the response of recharge. Interestingly, significant hydrogeologic predictors varied according to the ecoregion, with hydraulic conductivity playing an important role in ecoregions where LAI-driven ET differences are mostly partitioned to recharge, and calibrated runoff parameters playing an important role where there are major tile drain effects.

To assess drivers of the vegetation dynamics that affect recharge, we examined inter-annual climatic factors and found that large springtime temperature anomalies are responsible for the large springtime normalized LAI anomalies and recharge differences. This reflects the well-known relationship between temperature and leaf-emergence and cropping schedules in higher latitudes. Temperature anomalies do not seem to have a similar control on fall-time leaf senescence and harvest schedules, and precipitation appears to have a minor influence on LAI anomalies and recharge responses to vegetation dynamics, even in drier regions and seasons.

Our assessment of recharge responses to vegetation dynamics have important implications for recharge predictions under climate change, because it reveals that vegetation responses to climate should be accounted for. We showed that across a precipitation gradient, vegetation leaf-out (including crop schedules) across Minnesota is highly sensitive to springtime temperature anomalies, and this phenological response can trigger notable changes in ET and subsequently recharge. Although the high sensitivity of vegetation and recharge to temperature suggests that precipitation may be a less important factor at higher mid-latitudes, the response of recharge differed along the state's precipitation gradient. In the drier Great Plains to the west, springtime normalized LAI anomalies and corresponding recharge differences were greater than in the wetter eastern part of the state, despite having similar magnitude temperature anomalies, suggesting greater ecohydrological vulnerability to temperature anomalies with lower precipitation. Overall, our findings indicate that regardless of uncertain future projections of precipitation, projected temperature increases will extend growing seasons across Minnesota, which will exert a decreasing effect on springtime recharge that will be stronger in drier regions of the state, as well as where hydraulic conductivity is higher and the effects of tile drains are more minor (see Table 3).

Funding

Funding for this work came from the Legislative-Citizen Commission on Minnesota Resources (the Environment and Natural Resources Trust Fund (ENRTF), M.L. 2016, Chp. 186, Section 2, Subd. 03f) and from the National Science Foundation (EAR-1724781).

CRediT authorship contribution statement

Harsh Anurag: Methodology, Software, Formal analysis, Investigation, Writing - original draft, Visualization. **G.H. Crystal Ng:** Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Robert Tipping:** Conceptualization, Supervision, Funding acquisition. **Kathy Tokos:** Software, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors greatly appreciate assistance from Dr. Stefan Liess (University of Minnesota) to set up CLM on the University of Minnesota's computing cluster, and assistance from Dr. Shaoqing Liu (University of Minnesota) to implement CLM and the Ensemble Kalman Filter. We thank Erik Smith (USGS) for providing helpful details about his previous Minnesota recharge research (Smith and Westenbroek, 2015). The authors also acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing computing resources for this work.http://www.msi.umn.edu

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.jhydrol.2021.126584.

References

Adams, R. (2016). Water-table elevation and depth to water table: Minnesota Department of Natural Resources, Minnesota Hydrogeology Atlas Series HG-03, report and 2 plates. Technical Report Minnesota Department of Natural Resources.

Anderson, J.L., Anderson, S.L., 1999. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Weather Rev. 127, 2741–2758 url:https://doi.org/10.1175/1520-0493(1999)127<2741:amciot>2.0.co;2.

Arora, V., 2002. Modeling vegetation as a dynamic component in soil-vegetation-atmosphere transfer schemes and hydrological models. Rev. Geophys. 40, 1006. https://doi.org/10.1029/2001RG000103.

Asner, G.P., Scurlock, J.M.O., Hicke, J.A., 2003. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. Glob. Ecol. Biogeogr. 12, 191–205. https://doi.org/10.1046/j.1466-822x.2003.00026.x.

Bondeau, A., Kicklighter, D.W., Kaduk, J., Intercomparison, T.P.O.T.P., 1999. Comparing global models of terrestrial net primary productivity (NPP): importance of vegetation structure on seasonal NPP estimates. Glob. Change Biol. 5, 35–45. https://doi.org/10.1046/j.1365-2486.1999.00005.x.

Bouraoui, F., Vachaud, G., Li, L.Z.X., Treut, H.L., Chen, T., 1999. Evaluation of the impact of climate changes on water storage and groundwater recharge at the watershed scale. Clim. Dyn. 15, 153–161. https://doi.org/10.1007/s003820050274.

Braswell, B.H., 1997. The response of global terrestrial ecosystems to interannual temperature variability. Science 278, 870–873. https://doi.org/10.1126/science.278.5339.870.

Carrassi, Alberto, Bocquet, Marc, Bertino, Laurent, Evensen, Geir, 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. Wiley Interdiscip. Rev. Clim. Change. https://doi.org/10.1002/wcc.535.

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C. W., Odgers, N.P., 2016. POLARIS: A 30-meter probabilistic soil series map of the contiguous united states. Geoderma 274, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Chang, M. (2012). Forest Hydrology. CRC Press. url:https://doi.org/10.1201/b13614. doi:10.1201/b13614.

Chen, Z., Grasby, S.E., Osadetz, K.G., 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. J. Hydrol. 260, 102–117. https://doi.org/10.1016/s0022-1694(01)00606-0.

Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. Water Resour. Res. 14, 601–604. https://doi.org/10.1029/wr014i004p00601.

Clark, D.B., Mercado, L.M., Sitch, S., Jones, C.D., Gedney, N., Best, M.J., Pryor, M., Rooney, G.G., Essery, R.L.H., Blyth, E., Boucher, O., Harding, R.J., Huntingford, C., Cox, P.M., 2011. The joint UK land environment simulator (JULES), model description – part 2: Carbon fluxes and vegetation dynamics. Geoscientific Model Dev. 4, 701–722. https://doi.org/10.5194/gmd-4-701-2011.

Clark, M.P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D.E., Woods, R.A., Freer, J. E., Gutmann, E.D., Wood, A.W., Brekke, L.D., Arnold, J.R., Gochis, D.J., Rasmussen, R.M., 2015. A unified approach for process-based hydrologic modeling: 1. modeling concept. Water Resour. Res. 51, 2498–2514. https://doi.org/10.1002/2015wr017198.

Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20, 682–690. https://doi.org/10.1029/wr020i006p00682.

Croley, T.E., Luukkonen, C.L., 2003. Potential effects of climate change on ground water in lansing, michigan. JAWRA J. Am. Water Resources Assoc. 39, 149–163.

Crosbie, R.S., McCallum, J.L., Walker, G.R., Chiew, F.H.S., 2010. Modelling climate-change impacts on groundwater recharge in the murray-darling basin, australia. Hydrogeol. J. 18, 1639–1656. https://doi.org/10.1007/s10040-010-0625-x.
Delin, G.N., Healy, R.W., Lorenz, D.L., Nimmo, J.R., 2007. Comparison of local- to regional-scale estimates of ground-water recharge in minnesota, USA. J. Hydrol. 334, 231–249. https://doi.org/10.1016/j.jhydrol.2006.10.010.
Döll, P., 2009. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. Environ. Res. Lett. 4, 035006 https://doi.org/10.1088/1748-9326/4/3/035006.
Druel, A., Peylin, P., Krinner, G., Ciais, P., Viovy, N., Peregon, A., Bastrikov, V., Kosykh, N., Mironycheva-Tokareva, N., 2017. Towards a more detailed representation of high-latitude vegetation in the global land surface model ORCHIDEE (ORC-HL-VEGv1.0). Geoscientific Model Dev. 10, 4693–4722. https://doi.org/10.5194/gmd-10-4693-2017.
Eagleson, P.S., 1978. Climate, soil, and vegetation: 1. introduction to water balance dynamics. Water Resour. Res. 14, 705–712. https://doi.org/10.1029/wr014i005p00705.
Eckhardt, K., Ulbrich, U., 2003. Potential impacts of climate change on groundwater recharge and streamflow in a central european low mountain range. J. Hydrol. 284, 244–252. https://doi.org/10.1016/j.jhydrol.2003.08.005.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143. https://doi.org/10.1029/94jc00572.
Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An overview of global leaf area index (LAI): Methods, products, validation, and applications. Rev. Geophys. 57, 739–799. https://doi.org/10.1029/2018rg000608.
Fatichi, S., Pappas, C., Ivanov, V.Y., 2015. Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale. WIREs Water 3, 327–368. https://doi.org/10.1002/wat2.1125.
Fernandez-Illescas, C.P., Porporato, A., Laio, F., Rodriguez-Iturbe, I., 2001. The ecohydrological role of soil texture in a water-limited ecosystem. Water Resour. Res. 37, 2863–2872. https://doi.org/10.1029/2000wr000121.
Fisher, R.A., Koven, C.D., Anderegg, W.R.L., Christoffersen, B.O., Dietze, M.C., Farrior, C. E., Holm, J.A., Hurtt, G.C., Knox, R.G., Lawrence, P.J., Lichstein, J.W., Longo, M., Matheny, A.M., Medvigy, D., Muller-Landau, H.C., Powell, T.L., Serbin, S.P., Sato, H., Shuman, J.K., Smith, B., Trugman, A.T., Viskari, T., Verbeeck, H., Weng, E., Xu, C., Xu, X., Zhang, T., Moorcroft, P.R., 2017. Vegetation demographics in earth system models: A review of progress and priorities. Glob. Change Biol. 24, 35–54. https://doi.org/10.1111/gcb.13910.
Ford, T.W., Quiring, S.M., 2013. Influence of MODIS-derived dynamic vegetation on VIC-simulated soil moisture in oklahoma. J. Hydrometeorol. 14, 1910–1921. https://doi.org/10.1175/jhm-d-13-037.1.
Gleeson, T., Wada, Y., Bierkens, M.F.P., van Beek, L.P.H., 2012. Water balance of global aquifers revealed by groundwater footprint. Nature 488, 197–200. https://doi.org/10.1038/nature11295.
Green, T.R., Bates, B.C., Charles, S.P., Fleming, P.M., 2007. Physically based simulation of potential effects of carbon dioxide-altered climates on groundwater recharge. Vadose Zone J. 6, 597–609. https://doi.org/10.2136/vzj2006.0099.
Gunderson, C.A., Edwards, N.T., Walker, A.V., O'Hara, K.H., Campion, C.M., Hanson, P. J., 2012. Forest phenology and a warmer climate - growing season extension in relation to climatic provenance. Glob. Change Biol. 18, 2008–2025. https://doi.org/10.1111/j.1365-2486.2011.02632.x.
Hanson, R.T., Dettinger, M.D., 2005. Groundwater/surface water responses to global climate simulations, santa clara - calleguas basin, ventura, california. J. Am. Water Resour. Assoc. 41, 517–536. https://doi.org/10.1111/j.1752-1688.2005.tb03752.x.
Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. Proceedings of the National Academy of Sciences, 114, 2842–2847. doi:10.1073/pnas.1614941114.
Herrera-Pantoja, M., Hiscock, K.M., 2007. The effects of climate change on potential groundwater recharge in great britain. Hydrol. Process. 22, 73–86. https://doi.org/10.1002/hyp.6620.
Huang, M., Ray, J., Hou, Z., Ren, H., Liu, Y., Swiler, L., 2016. On the applicability of surrogate-based markov chain monte carlo-bayesian inversion to the community land model: Case studies at flux tower sites. J. Geophys. Res.: Atmospheres 121, 7548–7563. https://doi.org/10.1002/2015jd024339.
Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J.-F., Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W.D., Hack, J.J., Kiehl, J., Marshall, S., 2013. The community earth system model: A framework for collaborative research. Bull. Am. Meteorol. Soc. 94, 1339–1360. https://doi.org/10.1175/bams-d-12-00121.1.
Hwang, T., Martin, K.L., Vose, J.M., Wear, D., Miles, B., Kim, Y., Band, L.E., 2018. Nonstationary hydrologic behavior in forested watersheds is mediated by climate-induced changes in growing season length and subsequent vegetation growth. Water Resour. Res. 54, 5359–5375. https://doi.org/10.1029/2017wr022279.
Jeong, S.-J., HO, C.-H., Gim, H.-J., & Brown, M.E. (2011). Phenology shifts at start vs. end of growing season in temperate vegetation over the northern hemisphere for the period 1982-2008. Global Change Biology, 17, 2385–2399. doi:10.1111/j.1365-2486.2011.02397.x.
Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6, 2001–2013 doi:10.5194 10.5194/bg-6-2001-2009.
Jyrkama, M.I., Sykes, J.F., 2007. The impact of climate change on spatially varying groundwater recharge in the grand river watershed (ontario). J. Hydrol. 338, 237–250. https://doi.org/10.1016/j.jhydrol.2007.02.036.
Kergoat, L., 1998. A model for hydrological equilibrium of leaf area index on a global scale. J. Hydrol. 212–213, 268–286. https://doi.org/10.1016/s0022-1694(98)00211-x.
Kim, J.H., Jackson, R.B., 2012. A global analysis of groundwater recharge for vegetation, climate, and soils. Vadose Zone J. 11 https://doi.org/10.2136/vzj2011.0021ra.
Kirshen, P.H., 2002. Potential impacts of global warming on groundwater in eastern massachusetts. Journal of Water Resources Planning and Management 128, 216–226. https://doi.org/10.1061/(asce)0733-9496(2002)128:3(216).
Laio, F., Tamea, S., Ridolfi, L., D'Odorico, P., Rodriguez-Iturbe, I., 2009. Ecohydrology of groundwater-dependent ecosystems: 1. stochastic water table dynamics. Water Resour. Res. 45 https://doi.org/10.1029/2008wr007292.
Lawrence, D.M., Slater, A.G., 2007. Incorporating organic soil into a global climate model. Clim. Dyn. 30, 145–160. https://doi.org/10.1007/s00382-007-0278-1.
Lawrence, P.J., Chase, T.N., 2007. Representing a new MODIS consistent land surface in the community land model (CLM 3.0). J. Geophys. Res. 112 https://doi.org/10.1029/2006jg000168.
Le Maitre, D.C., Scott, D.F., Colvin, C., 1999. A review of information on interactions between vegetation and groundwater. Water SA 25, 137–152.
Levis, S., Bonan, G., Vertenstein, M., Oleson, K., 2004. The Community Land Model's Dynamic Global Vegetation Model (CLM-DGVM): Technical description and user's guide. Technical Report.
Liu, S., Ng, G.-H.C., 2019. A data-conditioned stochastic parameterization of temporal plant trait variability in an ecohydrological model and the potential for plasticity. Agric. For. Meteorol. 274, 184–194. https://doi.org/10.1016/j.agrformet.2019.05.005.
Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resour. Res. 43 https://doi.org/10.1029/2006wr005756.
Loheide, S.P., Gorelick, S.M., 2007. Riparian hydroecology: A coupled model of the observed interactions between groundwater flow and meadow vegetation patterning. Water Resour. Res. 43 https://doi.org/10.1029/2006wr005233.
Los, S.O., 2013. Analysis of trends in fused AVHRR and MODIS NDVI data for 1982–2006: Indication for a CO 2 fertilization effect in global vegetation. Global Biogeochem. Cycles 27, 318–330. https://doi.org/10.1002/gbc.20027.
Mao, J., Shi, X., Thornton, P.E., Piao, S., Wang, X., 2012. Causes of spring vegetation growth trends in the northern mid–high latitudes from 1982 to 2004. Environ. Res. Lett. 7, 014010 https://doi.org/10.1088/1748-9326/7/1/014010.
Meixner, T., Manning, A.H., Stonestrom, D.A., Allen, D.M., Ajami, H., Blasch, K.W., Brookfield, A.E., Castro, C.L., Clark, J.F., Gochis, D.J., Flint, A.L., Neff, K.L., Niraula, R., Rodell, M., Scanlon, B.R., Singha, K., Walvoord, M.A., 2016. Implications of projected climate change for groundwater recharge in the western united states. J. Hydrol. 534, 124–138. https://doi.org/10.1016/j.jhydrol.2015.12.027.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North american regional reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360. https://doi.org/10.1175/bams-87-3-343.
MNDNR (1993). Regional Hydrogeologic Assessment: Anoka Sand Plain; RHA-1. Technical Report Minnesota Department of Natural Resources.
Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state–parameter estimation of hydrological models using ensemble kalman filter. Adv. Water Resour. 28, 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002.
Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111, 519–536. https://doi.org/10.1016/j.rse.2007.04.015.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.
Ng, G.-H.C., Bedford, D.R., Miller, D.M., 2014. A mechanistic modeling and data assimilation framework for mojave desert ecohydrology. Water Resour. Res. 50, 4662–4685. https://doi.org/10.1002/2014wr015281.
Ng, G.-H.C., McLaughlin, D., Entekhabi, D., Scanlon, B., 2009. Using data assimilation to identify diffuse recharge mechanisms from chemical and physical data in the unsaturated zone. Water Resour. Res. 45 https://doi.org/10.1029/2009wr007831.
Ng, G.-H.C., McLaughlin, D., Entekhabi, D., Scanlon, B.R., 2010. Probabilistic analysis of the effects of climate change on groundwater recharge. Water Resour. Res. 46 https://doi.org/10.1029/2009wr007904.
Niu, G.-Y., Yang, Z.-L., Dickinson, R.E., Gulden, L.E., Su, H., 2007. Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. J. Geophys. Res. 112 https://doi.org/10.1029/2006jd007522.
Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M. et al. (2013). Technical Description of Version 4.5 of the Community Land Model (CLM). Technical Report NCAR Technical Note NCAR/TN-503+ STR, Boulder, Colorado.
Pearson, S., Falteisek, J., Berg, J., 2012. Minnesota Groundwater Level Monitoring Network - Guidance Document for Network Development. Technical Report.
Peck, A., Williamson, D., 1987. Effects of forest clearing on groundwater. J. Hydrol. 94, 47–65. https://doi.org/10.1016/0022-1694(87)90032-1.
Petheram, C., Walker, G., Grayson, R., Thierfelder, T., Zhang, L., 2002. Towards a framework for predicting impacts of land-use on recharge: 1. a review of recharge studies in australia. Soil Res. 40, 397. https://doi.org/10.1071/sr00057.

Piao, S., Mohammat, A., Fang, J., Cai, Q., Feng, J., 2006. NDVI-based increase in growth of temperate grasslands and its responses to climate changes in china. Global Environ. Change 16, 340–348. https://doi.org/10.1016/j.gloenvcha.2006.02.002.

Porporato, A., Laio, F., Ridolfi, L., Rodriguez-Iturbe, I., 2001. Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress. Adv. Water Resour. 24, 725–744. https://doi.org/10.1016/s0309-1708(01)00006-9.

Quetin, G.R., Swann, A.L.S., 2018. Sensitivity of leaf area to interannual climate variation as a diagnostic of ecosystem function in CMIP5 carbon cycle models. J. Climate 31, 8607–8625. https://doi.org/10.1175/jcli-d-17-0580.1.

Ray, J., Hou, Z., Huang, M., Sargsyan, K., Swiler, L., 2015. Bayesian calibration of the community land model using surrogates. SIAM/ASA J. Uncertainty Quantification 3, 199–233. https://doi.org/10.1137/140957998.

Reichle, Rolf H., 2008. Data assimilation methods in the Earth sciences. Adv. Water Resour. https://doi.org/10.1016/j.advwatres.2008.01.001.

Ringeval, B., Decharme, B., Piao, S.L., Ciais, P., Papa, F., de Noblet-Ducoudré, N., Prigent, C., Friedlingstein, P., Gouttevin, I., Koven, C., Ducharne, A., 2012. Modelling sub-grid wetland in the ORCHIDEE global land surface model: evaluation against river discharges and remotely sensed data. Geoscientific Model Dev. 5, 941–962. https://doi.org/10.5194/gmd-5-941-2012.

Rodriguez-Iturbe, I., 2000. Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies. Water Resour. Res. 36, 3–9. https://doi.org/10.1029/1999wr900210.

Rodriguez-Iturbe, I., D'Odorico, P., Porporato, A., Ridolfi, L., 1999. On the spatial and temporal links between vegetation, climate, and soil moisture. Water Resour. Res. 35, 3709–3722. https://doi.org/10.1029/1999wr900255.

Scanlon, B.R., Keese, K.E., Flint, A.L., Flint, L.E., Gaye, C.B., Edmunds, W.M., Simmers, I., 2006. Global synthesis of groundwater recharge in semiarid and arid regions. Hydrol. Process. 20, 3335–3370. https://doi.org/10.1002/hyp.6335.

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. J. Am. Water Resour. Assoc. 49, 577–591 doi:10.1111 10.1111/jawr.12057.

Smith, E.A., Gillette, T., Blann, K., Coburn, M., Hoppie, B., & Rhees, S., 2018. Drain tiles and groundwater resources: Understanding the relations. Technical Report Minnesota Ground Water Association.

Smith, E.A., & Westenbroek, S.M., 2015. Potential groundwater recharge for the state of minnesota using the soil-water-balance model, 1996–2010. url:https://doi.org/10.3133/sir20155038. doi:10.3133/sir20155038.

Soil Survey Staff (2020). Web soil survey. url:https://websoilsurvey.nrcs.usda.gov/natural Resources Conservation Service, United States Department of Agriculture.

Sun, Leqiang, Seidou, Ousmane, Nistor, Ioan, Liu, Kailei, 2016. Review of the Kalman-type hydrological data assimilation. Hydrolog. Sci. J. https://doi.org/10.1080/02626667.2015.1127376.

Tang, Q., Vivoni, E.R., Muñoz-Arriola, F., Lettenmaier, D.P., 2012. Predictability of evapotranspiration patterns using remotely sensed vegetation dynamics during the north american monsoon. J. Hydrometeorol. 13, 103–121. https://doi.org/10.1175/jhm-d-11-032.1.

Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2012. Ground water and climate change. Nature. Clim. Change 3, 322–329. https://doi.org/10.1038/nclimate1744.

Tesemma, Z., Wei, Y., Peel, M., Western, A., 2015. The effect of year-to-year variability of leaf area index on variable infiltration capacity model performance and simulation of runoff. Adv. Water Resour. 83, 310–322. https://doi.org/10.1016/j.advwatres.2015.07.002.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., & Cook, R.B. (2014). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Technical Report Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States).

Twin Cities Metropolitan Council (2007). Water supply planning in the Twin Cities metropolitan area. Technical Report Metropolitan Council.

Vivoni, E.R., 2012. Diagnosing seasonal vegetation impacts on evapotranspiration and its partitioning at the catchment scale during SMEX04–NAME. Journal of Hydrometeorology 13, 1631–1638. https://doi.org/10.1175/jhm-d-11-0131.1.

Wang, X., Piao, S., Ciais, P., Li, J., Friedlingstein, P., Koven, C., & Chen, A. (2011). Spring temperature change and its implication in the change of vegetation growth in north america from 1982 to 2006. Proceedings of the National Academy of Sciences, 108, 1240–1245. url:https://doi.org/10.1073 doi:10.1073/pnas.1014425108.

Wegehenkel, M., 2009. Modeling of vegetation dynamics in hydrological models for the assessment of the effects of climate change on evapotranspiration and groundwater recharge. Adv. Geosci. 21, 109–115. https://doi.org/10.5194/adgeo-21-109-2009.

White, D. (2020). Ecological Regions of Minnesota: Level III and IV maps and descriptions. 22 pages text, 69 pages appendices. Technical Report.

Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., Zhao, W., 2015. Time-lag effects of global vegetation responses to climate change. Glob. Change Biol. 21, 3520–3531. https://doi.org/10.1111/gcb.12945.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., Song, J., 2014. Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. IEEE Trans. Geosci. Remote Sens. 52, 209–223. https://doi.org/10.1109/tgrs.2013.2237780.

Xu, L., Myneni, R.B., III, F.S.C., Callaghan, T.V., Pinzon, J.E., Tucker, C.J., Zhu, Z., Bi, J., Ciais, P., Tø;mmervik, H., Euskirchen, E.S., Forbes, B.C., Piao, S.L., Anderson, B.T., Ganguly, S., Nemani, R.R., Goetz, S.J., Beck, P.S.A., Bunn, A.G., Cao, C., & Stroeve, J. C. (2013). Temperature and vegetation seasonality diminishment over northern lands. Nature Climate Change, 3, 581–586. url:https://doi.org/10.1038/nclimate1836. doi:10.1038/nclimate1836.

Zeng, X., Decker, M., 2009. Improving the numerical solution of soil moisture–based richards equation for land models with a deep or shallow water table. J. Hydrometeorol. 10, 308–319. https://doi.org/10.1175/2008jhm1011.1.

Zeng, Y., Xie, Z., Liu, S., Xie, J., Jia, B., Qin, P., Gao, J., 2018. Global land surface modeling including lateral groundwater flow. J. Adv. Modeling Earth Syst. 10, 1882–1900. https://doi.org/10.1029/2018ms001304.

Zhang, Y., Chiew, F.H.S., Zhang, L., Li, H., 2009. Use of remotely sensed actual evapotranspiration to improve rainfall–runoff modeling in southeast australia. Journal of Hydrometeorology 10, 969–980. https://doi.org/10.1175/2009jhm1061.1.

Zheng, G., Moskal, L.M., 2009. Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. Sensors 9, 2719–2745. https://doi.org/10.3390/s90402719.

Zhu, Z., Piao, S., Myneni, R.B., Huang, M., Zeng, Z., Canadell, J.G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T.A.M., Stocker, B.D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., Zeng, N., 2016. Greening of the earth and its drivers. Nature Clim. Change 6, 791–795. https://doi.org/10.1038/nclimate3004.