Predicting patterns of solar energy buildout to identify opportunities for biodiversity conservation

Michael J. Evans a,b Kumar Mainali a, c Rachel Soobitsky (SSAI) Emily Mills a Susan Minnemeyer a
a Conservation Innovation Center, Chesapeake Conservancy, 716 Giddings Ave., Suite 42, Annapolis, MD
b Environmental Science and Policy Dept., George Mason University, 4400 University Dr., Fairfax, VA
c Department of Biology, University of Maryland, 4094 Campus Dr., College Park, MD 20742, USA

Abstract The construction of solar energy facilities can have positive or negative impacts on biodiversity depending on siting and associated land use transitions. We identified drivers of solar siting and quantified patterns of buildout in states surrounding the Chesapeake Bay watershed – a biodiversity hotspot with numerous ecosystem services. Using a convolutional neural network, we mapped the footprints of ground-mounted solar arrays present in satellite imagery annually from 2017 – 2021 in Delaware, Maryland, Pennsylvania, New York, Virginia, and West Virginia. As of 2021, we identified 958 solar arrays covering 523.2 km^2 built primarily on previously cultivated land, while avoiding natural landcover. We fit a binomial-Weibull model to these solar timeseries data in a hierarchical, Bayesian framework to quantify the relationship between geospatial covariates and rate of solar development. Solar array construction rate increased in cultivated areas, areas of lower agricultural suitability, lower slope, lower forest cover, lower biodiversity protection, and greater distances from roads. We also estimated changes in the rate of solar construction over time and found differences among states: acceleration in Virginia and deceleration in New York. We used parameter estimates to map the relative likelihood of future solar development across the study area. This methodology can be used to anticipate where solar is likely to be built in different landscapes and how these patterns align with conservation goals. Around the Chesapeake Bay watershed, the selection of lower quality agricultural areas for solar energy minimizes removal of important habitat and provides opportunities for native plant and pollinator restoration.

Keywords: Artificial Intelligence; Bayesian; Biodiversity; Conservation; Land Use; Remote Sensing; Renewable energy

1. Introduction

Climate change is one of the primary human drivers of biodiversity loss and the degradation of natural ecosystems (Díaz et al., 2019). Unless greenhouse gas emissions are drastically reduced millions of species face extinction (IPCC, 2022), many in the coming decades (Trisos et al., 2020). A transition from fossil fuels to renewable energy sources, especially wind and solar, comprise a prominent path to reduce carbon emissions (Rogelj et al., 2018; United States Government, 2021). Such an energy transition is therefore critical to conserve biodiversity and safeguard ecosystem services on which human society depends. However, land use change is also a major driver of habitat and biodiversity loss (Díaz et al., 2019), and the buildout of additional energy production facilities that fragment or destroy natural landcover may directly harm species and reduce ecosystem services, even if they mitigate climate change. The question of how to expand renewable energy production in a sustainable manner is therefore central to efforts to conserve biodiversity.

Approaches to synergize renewable energy buildout and biodiversity conservation are needed as countries around the world implement renewable energy goals. Solar energy, which includes both concentrated solar power and photovoltaic arrays, has been a leading means by which countries meet these goals, proliferating globally (Kruitwagen et al., 2021). The United States has made an aggressive commitment to 100% renewable energy by 2030 (United States Government, 2021), and already solar comprised 30% of additions to electricity generation capacity from 2015 – 2019. In 2020 solar accounted for 40% of U.S. additions, eclipsing all non-renewable energy sources (Feldman et al., 2022). By 2050, solar energy generation is expected to provide 22% of electricity in the U.S., with the great majority coming from photovoltaic arrays (U.S. Energy Information Administration, 2022). Electricity generation methods vary in area of

land required to produce a unit of energy; a measure known as land use intensity. Photovoltaic arrays are less land use intensive than other renewables like wind, hydropower, and biofuels, but more intensive than high-emission methods like coal and natural gas (McDonald et al. 2009).

Higher land use requirements add to the anthropogenic demands placed on limited land resources, which can threaten species with extinction due to habitat loss and fragmentation (Díaz et al., 2019; Fahrig, 2003). If the construction of solar energy involves the removal of natural land cover that provides habitat for wildlife, this transition can have immediate, negative environmental impacts (Hernandez et al., 2014; Tsoutsos et al., 2005). The addition of impervious surfaces may increase runoff (Pisinaras et al., 2014), and photovoltaic arrays can alter microclimates by creating heat islands (Barron-Gafford et al., 2016). Taken together, the effect of these processes may result in reduced species richness and diversity within the bounds of solar energy facilities, with a shift towards fewer rare species and more non-native species (Lovich and Ennen, 2011; Visser et al., 2019). Even when solar construction occurs without removing natural landcover, it can compete with other anthropogenic land uses (Sacchelli et al., 2016; Wu et al., 2020). If these uses are displaced to native ecosystems the effects on biodiversity may be the same.

On the other hand, the construction of solar arrays may benefit biodiversity and native ecosystems depending on siting and prior land use. Co-locating solar energy with existing infrastructure or constructing facilities on degraded lands, when possible, is considered optimal from a conservation perspective (Macknick et al., 2013; Moore-O'Leary et al., 2017). When construction on natural lands in unavoidable, locating solar energy facilities in areas of low importance to wildlife can minimize negative impacts on biodiversity (Hernandez et al., 2019, 2015b). Furthermore, land at solar energy facilities can be managed to support native species,

including pollinators (Graham et al., 2021), and biodiversity outcomes may be positive if such practices are implemented (Sinha et al., 2018). Thus, it is possible to develop solar energy capacity in sustainable ways that are synergistic with natural ecosystems. While there has been much research providing pathways for sustainable development, little evidence exists that these tools and models have informed siting (Agha et al., 2020). To assess whether solar is being developed in a sustainable manner, and to maximize the potential to implement sustainable approaches, two fundamental questions must be answered: where has solar been developed, and where is it likely to be developed in the future?

In this study we analyze spatiotemporal patterns of solar energy development in six states overlapping the Chesapeake Bay watershed (hereafter 'study area'), using a new dataset containing footprints of ground-mounted solar arrays built from 2017 to 2021. The Chesapeake Bay is the largest estuarine ecosystem in the United States, and the third largest in the world, supporting biodiversity and ecosystem services of national and global importance (Kemp et al., 2005; CBF, 2014). The Chesapeake Bay has also been heavily degraded by a myriad different land uses within the watershed as well as activities within the bay itself. As such, it remains the focus of substantial restoration and conservation efforts (Chesapeake Bay Program, 2014). While prior research on the development and impact of solar energy have focused on arid landscapes like the Southwest U.S. (Agha et al., 2020; Cameron et al., 2012; Parker et al., 2018), globally solar energy is being built in a variety of ecosystems including ones like those around the Chesapeake Bay. We had three objectives: first, to produce a current map of solar arrays within the study area; second, to develop a model of solar development probability that can be used to anticipate places that are most likely to be developed for solar energy; third, to quantify the land cover transitions related to past solar development within the study area. In this third objective,

we also analyze Gap Analysis Project (GAP) species habitat richness data (Gergely et al., 2019). While other measures of biodiversity exist (e.g., Global Biodiversity Information Facility and iNaturalist occurrence data), the primary focus of this study is on land cover because it was measured consistently across the study area, and more likely directly affects siting decisions. Together, our results provide a mechanistic understanding of the solar siting process that can help quantify impacts to biodiversity and anticipate how continued solar development aligns with conservation and restoration goals (Chesapeake Bay Program, 2014).

2. Materials and Methods

2.1 Study area

Our study area comprised the entirety of the states of Delaware, Maryland, New York, Pennsylvania, Virginia, and West Virginia (Fig. 1). Each of these states overlap a part of the Chesapeake Bay watershed. Patterns of land use and land cover within parts of these states affect the Chesapeake ecosystem, and we analyze the entirety of each state for completeness and intra-state homogeneity.

2.2 Data

We mapped ground-mounted photovoltaic arrays (hereafter 'solar arrays') that occurred within the study area in each year from 2017 to 2021 (hereafter the 'study period'). Polygons representing solar arrays were generated by a U-Net model (Ronneberger et al., 2015) trained to delineate these features from Sentinel-2 (Drusch et al., 2012) satellite imagery (Evans et al., 2020). As input to this model, we created composite Sentinel-2 images for each year in the study period. In each year, we selected images collected between May and September, masked cloudy pixels, and created a median mosaic containing the blue, green, red, near-infrared, short-wave infrared 1, and short-wave infrared 2 bands. These annual image composites were broken into

overlapping 384x384 pixel chips, with each chip sharing 64 pixels with any adjacent chips. Chips were ingested by the trained U-Net model, which output per-pixel probabilities of a solar array. Overlapping buffers were trimmed from each pixel, resulting in a set of spatially contiguous 256x256 chips comprising an output solar probability image for the study area. We converted this probability image to solar array polygons using a 0.9 probability threshold and inspected each output polygon against current Sentinel-2 and Google Earth imagery to ensure it corresponded to a true solar array. Likewise, we used Energy Information Administration (EIA) point data to identify omissions. Sentinel-2 image processing, chip generation, solar array postprocessing and validation were performed using the Google Earth Engine (Gorelick et al., 2017) platform. U-Net model predictions were produced using Microsoft Azure Machine Learning Studio.

We quantified the types of land cover present at solar arrays prior to their construction using 30 m National Land Cover Dataset rasters (NLCD; Dewitz and U.S. Geological Survey, 2021). Within each solar array polygon present in 2021, we counted the number of pixels that fell into each land cover class prior to the array being constructed. For this tabulation, we used NLCD 2016 data for all solar arrays detected prior to 2020, and NLCD 2019 data for all solar arrays detected in 2020 or 2021. We converted these counts to the area (km^2) of each land cover class on which solar arrays were built.

Throughout the study area, we generated randomly located, rectangular polygons used to sample landscape covariates at locations where solar arrays had not been built. First, we generated random points in a 5:1 ratio with observed solar arrays. These points served as centroids of rectangles representing hypothetical arrays. Each rectangle had a 2:1 side ratio, with the length of the shortest side determined by the empirical distribution of the areas of mapped

solar arrays. For each observed solar array polygon, we calculated the smaller side of a 2:1 rectangle having the same area as that of the polygon. We used the mean and standard deviation of these measures to define a normal distribution, from which we randomly sampled the length of the shorter side of the random rectangles. Finally, we eliminated any polygons overlapping water. We refer to these polygons as simulated arrays.

At each observed and simulated array, we quantified potential biodiversity using 30 m GAP species habitat richness datasets, which count the number of species for which a given pixel is suitable habitat (Gergely et al., 2019). We summed richness among the four available taxonomic groups (amphibians, birds, mammals, and reptiles) and calculated mean per-pixel richness within each solar array polygon present in 2021. We also extracted a vector of landscape covariates that we hypothesized might be associated with the likelihood of solar array construction (Table 1). We generated random arrays and sampled geospatial covariates using the *Geopandas* (Jordahl et al., 2020), *Shapely* (Gillies, 2007), *rasterstats* (Perry, 2013), *Rasterio* (Gillies, 2013) and *earthengine-api* packages for Python 3.7.

2.2.Spatiotemporal patterns

Our first objective was to describe temporal and spatial patterns in the buildout of solar energy in the study area. We measured buildout as the area (km^2) of ground-mounted solar arrays. For each state, in each year, we calculated the total area of solar arrays, and the proportion of the respective state arrays comprised. We report the number of arrays detected, but our analyses focus on array area because the boundaries between individual arrays can be subjective and difficult to distinguish in 10 m resolution imagery, and subject to the choice of probability threshold used to convert U-Net probability outputs to solar array polygons. Therefore, area is a more reliable measure of buildout than array counts. We quantified patterns

of spatial clustering using the distribution of nearest neighbor indices (Clark and Evans, 1954) of arrays within each state.

To test for differences in mean array size and clustering among states we used one-way ANOVA models with state as a random effect. We modeled both variables as outcomes of a gaussian distribution, using a log link relating the expected mean to indicator variables per state. To estimate per-state rates of solar development over time, we fit an ANCOVA model to the proportion of states' area built out as solar in each year ($y_{s(i)}$). We refer to this response as 'proportional buildout'. We modeled mean proportional buildout as a function of year with a Gaussian error distribution and identity link. We fit a similar ANCOVA model of mean species habitat richness (n_{si}) within arrays as a function of whether arrays were real or simulated using a Poisson error distribution and log link:

$$y_{s(i)} \sim N(\mu_s, \sigma)$$

$$n_{s(i)} \sim Poisson(\lambda_s)$$

$$\mu_s = \alpha_{1s} + \Delta_s * year_i$$

$$\log(\lambda_s) = \alpha_{2s} + \beta_s * array_i$$

Both models included random effects per state on the intercepts (α_s) and slope parameters (Δ_s, β_s). Slope parameters estimated state-specific rates of proportional buildout per year (Δ_s), and differences in mean richness between real and simulated arrays (β_s). We refer to this parameter as the 'biodiversity selection.' We fit ANOVA and ANCOVA models in a Bayesian framework using JAGS software, which implements Markov Chain Monte Carlo (MCMC) estimation of parameter posterior distributions via a Gibbs sampling algorithm (Plummer, 2017). We fit models in JAGS using the *jagsUI* (Kellner and Meredith, 2021) package for R (R Core Team, 2021). We sampled posterior distributions from 3 MCMC chains of 10,000 iterations,

discarding the first 2,000 iterations as burn-in and every other sample thereafter. We used the Rhat statistic (which measures agreement between MCMC chains) < 1.1 to indicate algorithm convergence on the true posterior distribution for each estimated variable (Gelman and Rubin, 1992).

We specified uninformative prior distributions for the means (μ_α, μ_Δ) ~ *N(0, 10,000)* and standard deviations (σ_α, σ_Δ) ~ *U(0, 50)* of the normal distributions from which per-state random variables were drawn. For the ANOVA model, we report the mean and standard deviation of the posterior distributions of state-specific array size. We report the 95% credible interval of the posterior distributions of state-specific rate of proportional buildout (Δ_s) and state-specific biodiversity selection (β_s), and whether these intervals overlapped zero. Positive or negative intervals excluding zero would indicate selection for areas of high or low biodiversity importance, respectively. We tested for pairwise differences between states by measuring the overlap of posterior distributions. We considered an overlap of $p \leq 0.05$ to indicate a true difference.

We also summarized the types of land cover that was converted to solar arrays using all arrays present in 2021 and tested for selection and avoidance of specific land cover types. Selected classes were those for which the area converted to solar arrays was greater than expected, and avoided classes were those for which the area converted to solar arrays was less than expected. To test whether land cover classes were equally likely to be converted to solar arrays, we conducted Pearson's chi-squared test with the expected proportion of each NLCD land cover class equal to the proportional area of that class within the study area.

To estimate the strength of selection or avoidance of specific classes we compared the observed area of each NLCD class that was converted to solar arrays to an expected distribution

based on the availability of each class. We generated distribution of expected area converted by drawing 1,000 random samples from a multinomial distribution with number of classes equal to the number of land cover classes in the study area, total observations equal to the total area of solar arrays, and probability of each class equal to its proportion in the study area. We report the difference between the observed and mean expected converted areas as the effect size (km^2) of each class and the percentage of draws in which the observed area was greater or less than the random draw (p). We considered effects to differ significantly from zero if $p < 0.05$. This analysis was repeated within each state. In discussing these results, we group NLCD classes into three categories: Human Modified; Cultivated; and Natural to present land cover transitions in terms of conservation relevance.

In both analyses of selection for land cover, we calculated available area both including and excluding areas within lands with GAP Status 1 or 2. Natural land cover is protected from conversion in these areas and thus unavailable for solar construction. We present results excluding GAP 1 or 2 areas, and report when inclusion yielded different outcomes.

2.3.Predictive Modeling

To identify landscape characteristics associated with solar buildout, we used hierarchical Bayesian models to estimate the relationship between J landscape characteristics and the time until solar arrays were built at a given location during our study period. We modeled the time (years) to construction at a given site as a Weibull process conditional on the binomial outcome of a site being developed for solar. Conceptually, a Weibull distribution can be used to model time to event data, and accommodates increasing, decreasing, or constant rates of event occurrence. The joint binomial-Weibull model took the form:

$$s_i \sim Bernoulli(p_i)$$

$$d_i = s_i * I(y_i > ymax_i) + (1 - s_i)$$

$$y_i|s_i \sim Weibull(k_s, \lambda_i)T(l_i, \infty) \; if \; d_i = 0$$

$$y_i|s_i = NA \qquad if \; d_i = 1$$

$$k_s \sim Gamma(v_k, \theta_k)$$

$$\log(\lambda_i) = \log(\alpha_s) + \sum_{j=1}^{J} \beta_j x_{ji}$$

$$\alpha_s \sim Gamma(v_\alpha, \theta_\alpha)$$

Here, s_i was the binary measure of whether solar arrays were developed at a site, and y_i the years to solar development using 2016 as the first year. When solar arrays were not constructed by 2021 there was uncertainty as to whether an array may be constructed at that site in the future. Thus, we modeled y_i as a censored Weibull random variable. The time to construction at a site was censored at the maximum number of observation years ($ymax_i$) if arrays were not constructed by the end of the study period. We use an indicator variable d_i, to identify these sites, and treat the time to construction response as missing when $d_i = 1$.

We left-truncated the Weibull distribution of time to construction at a given site based on the minimum possible time at that site, l_i. In Virginia, Maryland, Pennsylvania, and Delaware cloud-free Sentinel-2 imagery was available in 5 years, whereas New York we were unable to map solar arrays in 2017 in New York. Therefore, l_i in New York was 2 (1 elsewhere) and $ymax_i$ was 4 in New York (5 elsewhere).

We defined the Weibull rate parameter, λ_i, as a function of a linear combination of J site covariates using a log link function. This linear predictor had a random intercept per state drawn from a Gamma distribution with shape v and rate θ. We also defined the Weibull shape parameter, k_s, as a random variable per state. The shape parameter indicates whether the

probability of an event is constant ($k = 1$), increasing ($k > 1$) or decreasing ($k < 1$) over time. Therefore, this formulation allowed us to estimate potential differences in the acceleration or deceleration in the rate of solar buildout among states.

We defined vaguely informative priors for all covariate coefficients $P(\beta_j) \sim N(\mu = 0, \sigma = 10{,}000)$ and the probability of solar array construction $P(p_i) \sim U(0,1)$. Instead of specifying priors on each of the shape (v) and rate (θ) parameters for Gamma distributions on k and α, we use moment matching to define these parameters in terms of their expected values and variances. For instance:

$$v_k = E[k]^2 / Var[k]$$

$$\theta_k = E[k] / Var[k]$$

We then specified prior distributions on the expected value and variance of the Gamma distribution for state-level shape parameters as $E[k] \sim Uniform(0, 6)$ and $Var[k] \sim Uniform(0, 1000)$, and the expected value and variance of the Gamma distribution for state-level intercepts for the linear predictor of rate parameters as $E[\alpha] \sim Uniform(0, 0.1)$ and $Var[\alpha] \sim Uniform(0, 1000)$.

We sampled posterior distributions of estimated parameters from 3 MCMC chains of 50,000 iterations, discarding the first 10,000 iterations as burn-in. We thinned chains by sampling every 2nd observation. We evaluated chain convergence using Rhat < 1.1. If after 50,000 iterations all variables had not converged, we performed additional fitting runs of 10,000 MCMC iterations of 3 chains until convergence was achieved.

We evaluated the fit of the full model by comparing residuals from datasets simulated under the model to those of the observed dataset. For each observation, i, at each MCMC iteration we used the current state of model parameters to calculate the expected response ($\overline{y_i}$) and generate a

simulated response from a Weibull distribution. We calculated residuals for the observed and simulated response values and quantified model fit as a chi-square statistic using all observations (Kéry and Royle, 2021). If the model fits the observed data we expect the observed and simulated chi square values at each observation to be equal. We present diagnostic plots of these fit statistics and report the proportion of observations for which the observed chi-square was greater than the simulated chi-square. This value should be close to 0.5 for a model that fits the data (Gelman et al., 1996).

Once we confirmed the model structure was appropriate for the data, we proceeded with a Bayesian variable selection procedure. We opted against an information theoretic model selection approach because of known issues with DIC (Spiegelhalter et al., 2002) when fitting hierarchical models (Lunn et al., 2013), and the impracticality of fitting all possible models in the model space. Therefore, we follow a variable selection procedure outlined in Kery & Royle (2016). After obtaining posterior estimates for all parameters from a full model exhibiting MCMC chain convergence, we re-fit the full model including binary indicator variables, w_j, for each beta parameter β_j in the linear predictor. Thus, the linear predictor in the variable selection model took the form:

$$\log(\alpha_s) + \sum_{j=1}^{J} w_j * \beta_j * X_{ji}$$

$$w_m \sim Bernoulli(0.5)$$

With this approach posterior model probabilities can be sensitive to the choice of prior distribution for the coefficients (Link and Barker, 2006). Therefore, we used Gibbs variable selection (Dellaportas et al., 2000) and specified joint priors for each coefficient following recommendations of (Tenan et al., 2014):

$$\beta_j \sim N(\mu_j, \sigma_j)$$

$$\mu_j = (1 - w_j) \cdot \mu_{jfull} + w_j \cdot 0$$

$$\sigma_j = (1 - w_j) \cdot \sigma_{jfull} + w_j \cdot 100$$

In this formulation the prior distribution of each coefficient, β_j, was equal to the posterior distribution from the full model $N(\bar{\beta}_{j\_full}, \sigma(\beta_{j\_full}))$, or an uninformative prior $N(0, 100)$ depending on the current value of the inclusion variable, w_j.

We calculate the posterior probability of variable inclusion, $P(w_j)$, for each potential predictor variable as the proportion of MCMC iterations for which the respective indicator w_j = 1. Using this procedure, we calculate model-averaged posterior means ($\overline{\beta}_j$) and standard deviations ($\sigma(\beta_j)$) for all covariate coefficients using MCMC samples when w_j = 1 (Kéry and Royle, 2021). We also report the overlap of model-averaged posterior distributions with zero as the probability that a coefficient was zero, $P(\beta_j = 0)$.

We use results of variable selection as evidence of the relative importance of different factors determining the location of solar buildout (O'Hara and Sillanpää, 2009). In addition to the mean and standard deviation of the posterior distributions of all coefficients, we report the mean ($\overline{\kappa}_s$) and standard deviations ($\sigma(k_s)$) of the posterior distributions of all per-state shape parameters. We used the posterior mean for all estimated coefficients and shape parameters to create two maps of the expected time to solar development at 30 m resolution across the CBW study area. In the first, each pixel contains the expected time to development value of a Weibull distribution parameterized using model-estimated coefficients and the observed covariate values at that pixel. In the second, each pixel contains the cumulative probability that time to development equals 10 under the same distribution.

All models were fit using JAGS software called from R v4.1.2 using the *jagsUI* package.

3. Results

3.1 Spatiotemporal patterns

We mapped 958 solar arrays covering 523.2 km^2 in 2021. Arrays were detected in all states except West Virginia (Fig. 1). EIA data indicated no omissions across the study area, and we excluded West Virginia from subsequent analyses. Mean array size was significantly larger in Virginia than all other states. The 95% credible intervals of biodiversity selection included zero in all states (Table 2), indicating the species habitat richness within solar arrays did not differ from that at random locations. Biodiversity selection did not differ between any state pairs ($p \geq 0.35$).

In all states, the area of solar arrays increased over time (Fig 2). In NY, cloudless images were not available in 2017, therefore we mapped and modeled annual solar buildout in NY each year from 2018 – 2021, and 2017 – 2021 in all other states. As of 2021, the most land had been developed for solar energy in Virginia, and the state with the greatest proportion of land developed for solar energy was Maryland (Table 2). The rate of proportional buildout in Virginia was greater than in Maryland ($p = 0.008$), and each of these states was greater than all others ($p = 0.00$). The rate of proportional buildout was lowest in Pennsylvania, with a 95% credible interval that included zero (Table 2).

Solar arrays were built on areas previously classified as 13 different NLCD land cover classes. Within the study area, the plurality of these classes were Cropland (156 km^2) and Pasture (74 km^2), accounting for 37.0% and 17.6% of the area on which solar arrays were built, respectively. While cropland was consistently the most frequently converted land cover class, the second most frequently converted class differed among states. In Delaware and Maryland,

Human Modified classes accounted for 3.2 km^2 (41.5%) and 26.7 km^2 (34.3 %) of the area on which solar arrays were built, respectively. In Virginia, Coniferous Forest accounted for 37 km^2 (18 %) of the area on which solar arrays were built.

The relative proportions of land cover classes on which solar arrays were built differed significantly from the available proportions of those classes within the study area (χ^2_{12} = 92,904, $p < 0.001$), and within each state ($\chi^2_{12} > 2428.22$, $p < 0.001$). Within the study area, Cultivated classes were most strongly selected for solar buildout while Natural classes were most strongly avoided (Fig. 3). This pattern was the same in each state (Fig. 3). Within each state, the only land cover classes on which the area of solar buildout did not differ from expected were Barren (*Effect* = -0.02 km^2, $p = 0.131$) in Delaware and Shrubland (*Effect* = 0.077 km^2, $p = 0.1048$) in Maryland. Including GAP Status 1 or 2 lands as available did not change the direction or significance of any effects.

3.2 Predictive analyses

The binomial-Weibull model of solar development fit time to solar construction data well ($p = 0.52$). Diagnostic plots of residuals as a function of observed values did not indicate any systematic deviation (Supplementary material). Rhat values were < 1.1 for all estimated parameters, indicating MCMC chain convergence.

Variable inclusion analysis indicated unequivocal support (i.e., $P(w_j) = 1.00$) for Cultivated, GAP Status, Latitude, and Road Distance as important predictors of λ_i (Table 3). Tree Cover, Slope, Farm Score, Open Space, Population, and Income were marginally supported ($0.00 < P(w_j) < 1.00$). Line Distance and Impervious were unequivocally not supported (i.e., $P(w_j) = 0.00$) as important predictors of λ_i (Table 3).

Model-averaged posterior estimates of coefficient values indicated positive relationships between λ_i and Cultivated, Farm Score, and Road Distance, and negative relationships between λ_i and Tree Cover, Slope, Latitude, and GAP Status. Model-averaged posterior estimates of coefficients for Open, Income, and Population included zero (Fig. 4). The mean of the posterior distributions for per-state Weibull shape parameters, $\overline{\kappa_s}$, were all greater than 1, except for New York (Fig. 4). Posterior shape parameter estimates were greatest for Virginia ($\overline{\kappa_{VA}}$ = 3.57, $\sigma(k_{VA}) = 0.18$) and lowest for New York ($\overline{\kappa_{NY}}$ = 0.81, $\sigma(k_{NY}) = 0.06$).

Spatial predictions of time to solar development reflected both variation in λ_i as a function of geographic covariates, and differences in k among states (Fig. 5). At a regional scale, areas with low expected time to development highlight hotspots of historical solar development, such as the eastern shore of Maryland and the Finger Lakes region of central New York (Fig. 5a). Cumulative probabilities illustrate where new hotpots may emerge in the future, such as Northwest Virginia (Fig. 5b). At a local scale, 30 m resolution predictions identify where new solar arrays are likely to be built first within states.

4. Discussion

Expanding renewable energy in a way that protects biodiversity is a key conservation challenge and understanding drivers of siting can help align buildout with conservation priorities. Solar energy facilities increased in number and area in states overlapping the Chesapeake Bay watershed, except West Virginia, from 2017 to 2021. We found strong evidence that solar arrays were built preferentially on previously cultivated lands, rarely replaced natural landcover, and were in areas of average habitat quality relative to available areas. The conversion of cropland and avoidance of natural landcover presents an opportunity to synergistically develop renewable energy and improve biodiversity, given the potential for solar energy facilities to be managed for

386 greater native species diversity and carbon sequestration relative to agriculture (Walston et al.,
387 2021). This pattern of buildout differed from those in the Southwest U.S. (Agha et al., 2020;
388 Cameron et al., 2012; Hernandez et al., 2015b; Parker et al., 2018), highlighting the need to
389 account for regional geographic, social, and political contexts when anticipating how solar
390 buildout will proceed. Our modeling approach explicitly accounted for variation in the rate of
391 solar energy buildout over time among administrative units (i.e., states), and we found
392 differences that may reflect variation in energy policies and regulations. Combining
393 sociopolitical and geographic variables, our results identified places where future solar buildout
394 is most likely, providing fundamental data needed to align current siting patterns with
395 conservation priorities in the ecological sensitive Chesapeake Bay region.
396 Solar development in the study area has thus far largely avoided important wildlife habitat.
397 Array sites avoided natural land cover like forest and wetlands, and did not select areas of higher
398 biodiversity importance. Even instances of ‘natural’ landcover conversion we detected may have
399 corresponded to cultivated lands. Several large solar array installations in Virginia were built in
400 areas of former timber harvest (i.e., coniferous forest), which are known to be confounded as
401 grassland, shrub, and pasture in NLCD classification (Wickham et al., 2017). Therefore, the
402 apparent selection for these classes may comprise a single land use transition from timber harvest
403 to solar energy. Given the critical role forest plays in carbon sequestration, an analysis
404 specifically identifying instances of this transition and assessing the net carbon effects could be
405 an important contribution to understanding the environmental impacts of solar siting. While
406 arrays were generally located outside places of high biodiversity importance, they also did not
407 select areas of low biodiversity importance. There may be greater potential to minimize the
408 ecological impact of renewable energy by prioritizing sites with low biodiversity importance

(Agha et al., 2020; Cameron et al., 2012). Additionally, the observed pattern of land use transition associated with solar buildout also provides additional means to protect and restore biodiversity.

In the study area, solar energy did not replace natural landcover because most facilities were built on cultivated land, which provides an opportunity to improve wildlife habitat. Cultivated land cover was the most frequently and most strongly selected type for solar array construction and was also among the most important predictors of solar buildout probability. While demonstrating that solar arrays have been preferentially sited on cultivated lands, these results cannot confirm whether this process constituted agricultural abandonment. The co-location of solar and agriculture is not only possible (Aroca-Delgado et al., 2018) but can improve yields for certain crops (Barron-Gafford et al., 2019). Importantly, the cultivated areas with less suitable farmland (i.e., higher scores) were more likely to be developed first. This was indicated by the positive relationship between Farm Score – a measure of agricultural suitability (1 = High, 5 = Low) - and the rate parameter of the Weibull model of time to solar development. Preferentially developing the least viable agricultural areas could make sense from an economic perspective and coincides with patterns of agricultural abandonment elsewhere (Gellrich and Zimmermann, 2007; Mottet et al., 2006). Geospatial model predictions indicate low-quality agricultural areas are most likely to be used for solar in the future.

Past and future conversion of cropland to solar arrays presents an opportunity to synergize renewable energy buildout and ecosystem restoration. From an ecological perspective, solar energy is optimally co-located on existing infrastructure or on previously degraded lands (Hernandez et al., 2019; Moore-O'Leary et al., 2017). While cropland is not considered degraded, it is ecologically less valuable than native landcover (Millennium Ecosystem

Assessment, 2005). Where former croplands have been abandoned, outcomes for biodiversity recovery may be mixed depending on land use history and degree of restoration management (Queiroz et al., 2014). In some agricultural areas native species diversity takes a long time to recover, if at all, and may require active restoration (Isbell et al., 2019). Land at solar energy facilities can be managed to support native plants and wildlife, and biodiversity may be increased at solar energy facilities (Macknick et al., 2013; Sinha et al., 2018; Walston., 2021). This includes local increases in pollinator abundance, diversity, and richness (Blaydes et al., 2021; Graham et al., 2021). Thus, the transition of cultivated areas to solar energy facilities could represent a net benefit to conservation if wildlife-friendly solar practices are implemented.

The macro patterns of solar development in the study area appear to differ from patterns in the Southwest U.S., where solar arrays have often been built near or within protected areas and replaced natural landcover (Hernandez et al., 2015b; Parker et al., 2018). In addition to the strong selection for agricultural lands, here solar array construction avoided natural landcover like deciduous forest and wetlands. Additionally, solar arrays were constructed away from protected areas, as indicated by the positive relationship between the rate parameter of the Weibull model of time to solar development and GAP Status. A variety of economic, political, and social factors determine where solar is ultimately built (Lawler et al., 2014). It is likely that the different geographic contexts between western states and this study area create different costs and benefits of avoiding natural landcover during solar siting. For instance, the Southwest U.S. contains large tracts of public lands managed for multiple use, including solar energy, and a larger proportion of natural landcover composed of shrubland and grassland. Thus, there may be less of a barrier to solar construction in natural ecosystems of the Southwest U.S. than in the CBW, where natural landcover and protected areas are largely forested. The differences in siting patterns illustrate the

need to account for regional drivers of solar development to understand its effect on biodiversity and the potential for synergy with conservation.

One potential driver of regional patterns are policies supporting solar development. Each of the five states in our study area have passed renewable energy portfolio standards that specify goals for the proportion of electricity generated within the state from renewable sources, but the amounts and timelines vary widely. This variation may explain differences in buildout rates among states. For example, Virginia had increased its standards in 2020 to include 100% renewable energy by 2040 (Virginia General Assembly, 2020). Conversely, West Virginia had no standard (West Virginia Legislature, 2015) and Pennsylvania had not updated their 2004 goal of 18% by 2020 (Pennsylvania General Assembly, 2004). Perhaps unsurprisingly, we detected no solar arrays in West Virginia, and Pennsylvania exhibited the lowest rate of expansion while buildout was most rapid in Virginia and Maryland. In Virginia, not only did solar buildout increase at the greatest rate since 2017, but our results indicate that rate is accelerating. The posterior estimate of the Weibull shape parameter for Virginia was much greater than one, which aligned with the observation that the greatest gain in solar array area in Virginia occurred between 2020 and 2021. Conversely, the posterior estimate of the Weibull shape parameter for New York was less than one, indicating buildout rates may be decelerating. While New York has committed to 70% renewable energy by 2030, legislation in 2019 specified a goal of 2,400 MW from offshore wind (Kaminsky, 2019), potentially indicating a shift away from solar. In addition to renewable energy portfolio standards avoidance of natural landcover during solar development may be facilitated by conservation policies (Maryland General Assembly, 1991) and tax incentives. Future research considering spatial variation in these policies at different scales can help clarify the role of sociopolitical drivers in solar siting that protects biodiversity.

5. Conclusion

The intentional siting of solar energy facilities offers a path to simultaneously shift the world's energy supply to renewable sources while protecting biodiversity. Spatial predictions of time to solar development, like those produced here, can help facilitate such siting by anticipating the future location of solar at two scales. Within the boundaries of administrative units among which differences in rate were estimated (i.e., states), relative differences in pixel values indicate where solar is most likely to be sited given some rate of future buildout. For instance, Virginia's goal of 40% renewable by 2030 would require approximately 31.7 km^2 of land, and we used a histogram of expected time to development values across the state to identify the most likely 31.7 km^2 to be converted. Combined with spatial data on energy demand and existing production, these data can be used to precisely anticipate where future solar buildout is likely to occur. Between administrative units, the effects of differences in the Weibull rate parameter are evident and allow us to anticipate new areas within the region are likely to experience solar development in the future. Around the Chesapeake Bay watershed, the likely locations of future solar construction identify potential ecological restoration opportunities. In combination with maps of existing habitat these data can help identify the most ecologically beneficial places for synergistic restoration and solar buildout. Predictions reflect the most likely locations for solar development under the continuation of current patterns. If these are sub-optimal for conservation, this advanced knowledge can be used to change policy to shift development to more desirable areas. While the results reported here apply to our study area, the approach to modeling solar buildout is appliable in any region where solar energy may come into conflict with conservation.

6. Data Availability

Georeferenced solar array polygon data used in analyses are available in an Open Science Framework repository (https://osf.io/vq7mt/). All code used in analysis of solar array data is available in an Open Science Framework repository (https://osf.io/9x6uv/).

References

Agha, M., Lovich, J.E., Ennen, J.R., Todd, B.D., 2020. Wind, sun, and wildlife: do wind and solar energy development 'short-circuit' conservation in the western United States? Environmental Research Letters 15, 075004. https://doi.org/10.1088/1748-9326/ab8846

Aroca-Delgado, R., Pérez-Alonso, J., Callejón-Ferre, Á., Velázquez-Martí, B., 2018. Compatibility between crops and solar panels: An overview from shading systems. Sustainability 10, 743. https://doi.org/10.3390/su10030743

Barron-Gafford, G.A., Minor, R.L., Allen, N.A., Cronin, A.D., Brooks, A.E., Pavao-Zuckerman, M.A., 2016. The Photovoltaic Heat Island Effect: Larger solar power plants increase local temperatures. Scientific Reports 6, 35070. https://doi.org/10.1038/srep35070

Barron-Gafford, G.A., Pavao-Zuckerman, M.A., Minor, R.L., Sutter, L.F., Barnett-Moreno, I., Blackett, D.T., Thompson, M., Dimond, K., Gerlak, A.K., Nabhan, G.P., Macknick, J.E., 2019. Agrivoltaics provide mutual benefits across the food–energy–water nexus in drylands. Nature Sustainability 2, 848–855. https://doi.org/10.1038/s41893-019-0364-5

Blaydes, H., Potts, S.G., Whyatt, J.D., Armstrong, A., 2021. Opportunities to enhance pollinator biodiversity in solar parks. Renewable and Sustainable Energy Reviews 145, 111065. https://doi.org/10.1016/J.RSER.2021.111065

Cameron, D.R., Cohen, B.S., Morrison, S.A., 2012. An Approach to Enhance the Conservation-Compatibility of Solar Energy Development. PLOS ONE 7, e38437. https://doi.org/10.1371/JOURNAL.PONE.0038437

Chesapeake Bay Program, 2014. Chesapeake Bay watershed agreement.

CBF, 2014. The economic benefits of cleaning up the Chesapeake Bay: A valuation of the natural benefits gained by implementing the Chesapeake clean water blueprint. Chesapeake Bay Foundation, Annapolis, MD. http://www.cbf.org/economicbenefits.

Clark, P.J., Evans, F.C., 1954. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. Ecology 35, 445–453. https://doi.org/10.2307/1931034

Dellaportas, P., Forster, J.J., Ntzoufras, I., 2000. Bayesian Variable Selection Using the Gibbs Sampler, in: Dey, D., Ghosh, S., Mallick, B. (Eds.), Generalized Linear Models: A Bayesian Perspective. CRC Press, New York, pp. 273–286.

Dewitz, J., U.S. Geological Survey, 2021. National Land Cover Database (NLCD) 2019 Products. U.S. Geological Survey data release. https://doi.org/https://doi.org/10.5066/P9KZCM54

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.-J., Visseren-Hamakers, I., Willis, K.J.,

Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science (1979) 366. https://doi.org/10.1126/science.aax3100

Drusch, M., del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Evans, M.J., Minich, T., Soobitsky, R., Mainali, K., 2020. A season independent U-net model for robust mapping of solar arrays using Sentinel-2 imagery. Preprints.org, 2020050345. https://doi.org/10.20944/preprints202005.0345.v2.

Fahrig, L., 2003. Effects of habitat fragmentation on biodiversity. Annual Review of Ecology, Evolution, and Systematics 34, 487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419

Feldman, D., Dummit, K., Zuboy, J., Heeter, J., Xu, K., Margolis, R., 2022. Winter 2021/2022 Solar Industry Update. National Renewable Energy Laboratory, Golden, CO.

Gellrich, M., Zimmermann, N.E., 2007. Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: A spatial statistical modelling approach. Landscape and Urban Planning 79, 65–76. https://doi.org/10.1016/J.LANDURBPLAN.2006.03.004

Gelman, A., Meng, X.-L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat Sin 6, 733–807.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7. https://doi.org/10.1214/ss/1177011136

Gillies, S., 2013. Rasterio: geospatial raster I/O for Python programmers.

Gillies, S., 2007. Shapely: manipulation and analysis of geometric objects.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Graham, M., Ates, S., Melathopoulos, A.P., Moldenke, A.R., DeBano, S.J., Best, L.R., Higgins, C.W., 2021. Partial shading by solar panels delays bloom, increases floral abundance during the late-season for pollinators in a dryland, agrivoltaic ecosystem. Scientific Reports 2021 11:1 11, 1–13. https://doi.org/10.1038/s41598-021-86756-4

Hernandez, R.R., Armstrong, A., Burney, J., Ryan, G., Moore-O'Leary, K., Diédhiou, I., Grodsky, S.M., Saul-Gershenz, L., Davis, R., Macknick, J., Mulvaney, D., Heath, G.A., Easter, S.B., Hoffacker, M.K., Allen, M.F., Kammen, D.M., 2019. Techno–ecological synergies of solar energy for global sustainability. Nature Sustainability 2, 560–568. https://doi.org/10.1038/s41893-019-0309-z

Hernandez, R.R., Easter, S.B., Murphy-Mariscal, M.L., Maestre, F.T., Tavassoli, M., Allen, E.B., Barrows, C.W., Belnap, J., Ochoa-Hueso, R., Ravi, S., Allen, M.F., 2014. Environmental impacts of utility-scale solar energy. Renewable and Sustainable Energy Reviews 29, 766–779. https://doi.org/10.1016/j.rser.2013.08.041

Hernandez, R.R., Hoffacker, M.K., Field, C.B., 2015a. Efficient use of land to meet sustainable energy needs. Nature Climate Change 5, 353–358. https://doi.org/10.1038/nclimate2556

Hernandez, R.R., Hoffacker, M.K., Murphy-Mariscal, M.L., Wu, G.C., Allen, M.F., 2015b. Solar energy development impacts on land cover change and protected areas. Proc Natl Acad Sci U S A 112, 13579–13584. https://doi.org/10.1073/PNAS.1517656112/SUPPL_FILE/PNAS.1517656112.ST05.DOCX

IPCC, 2022. Summary for policymakers, in: Portner, H.-O., Roberts, M., Tignor, M., Poloczanska, E.S., Mintenbeck, K., Alegria, A., Craig, M., Langsdorf, S., Loschke, S., Moller, V., Okem, A., Rama, B. (Eds.), Climate Change 2022: Impacts, Adaptation, and Vulnerability. Cambridge University Press.

Isbell, F., Tilman, D., Reich, P.B., Clark, A.T., 2019. Deficits of biodiversity and productivity linger a century after agricultural abandonment. Nature Ecology & Evolution 3, 1533–1538. https://doi.org/10.1038/s41559-019-1012-1

Jordahl, K., van den Bossche, J., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A.G., Farmer, C., Hjelle, G.A., Snow, A.D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L.J., Journois, M., WIlson, J., Greenhall, A., Holdgraf, C., Leblanc, F., 2020. geopandas. https://doi.org/10.5281/zenodo.3946761

Kaminsky, T., 2019. New York State climate leadership and community protection act. New York State Senate, Albany, NY.

Kellner, K., Meredith, M., 2021. jagsUI: a wrapper around “rjags” to streamline “JAGS” analyses.

Kemp, W.M., Boynton, W.R., Adolf, J.E., Boesch, D.F., Boicourt, W.C., Brush, G., Cornwell, J.C., Fisher, T.R., Gilbert, P.M., Hagy, J.D., 2005. Eutrophication of Chesapeake Bay: historical trends and ecological interactions. Marine Ecology Progress Series 303, 1-29.

Kéry, M., Royle, J.A., 2021. Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS, Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS. Elsevier. https://doi.org/10.1016/C2015-0-04070-9

Kruitwagen, L., Story, K.T., Friedrich, J., Byers, L., Skillman, S., Hepburn, C., 2021. A global inventory of photovoltaic solar energy generating units. Nature 598, 604–610. https://doi.org/10.1038/s41586-021-03957-7

613 Lawler, J.J., Lewis, D.J., Nelson, E., Plantinga, A.J., Polasky, S., Withey, J.C., Helmers, D.P.,
614 Martinuzzi, S., Penningtonh, D., Radeloff, V.C., 2014. Projected land-use change impacts on
615 ecosystem services in the United States. Proc Natl Acad Sci U S A 111, 7492–7497.
616 https://doi.org/10.1073/pnas.1405557111

617 Link, W.A., Barker, R.J., 2006. Model weights and the foundations of multimodel inference.
618 Ecology 87, 2626–2635.

619 Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. The BUGS Book: A
620 Practical Introduction to Bayesian Analysis. Chapman and Hall.

621 Lovich, J.E., Ennen, J.R., 2011. Wildlife conservation and solar energy development in the
622 desert southwest, United States. BioScience 61(12), 982-992.

623 Macknick, J., Beatty, B., Hill, G., 2013. Overview of Opportunities for Co-Location of Solar
624 Energy Technologies and Vegetation. Natural Renewable Energy Laboratory, Golden, CO.

625 Maryland General Assembly, 1991. Forest & Parks: Forest Conservation, Annotated Code of
626 Maryland.

627 McDonald, R.I., Fargione, J., Kiesecker, J., Miller, W.M., Powell, J., 2009. Energy sprawl or
628 energy efficiency: Climate policy impacts on natural habitat for the United States of
629 America. PLoS ONE 4(8), e6802. https://doi.org/10.1371/journal.pone.0006802

630 Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Biodiversity
631 Synthesis. World Resources Institute, Washington, DC.

632 Moore-O'Leary, K.A., Hernandez, R.R., Johnston, D.S., Abella, S.R., Tanner, K.E., Swanson,
633 A.C., Kreitler, J., Lovich, J.E., 2017. Sustainability of utility-scale solar energy – critical
634 ecological concepts. Frontiers in Ecology and the Environment 15, 385–394.
635 https://doi.org/10.1002/fee.1517

636 Mottet, A., Ladet, S., Coqué, N., Gibon, A., 2006. Agricultural land-use change and its drivers in
637 mountain landscapes: A case study in the Pyrenees. Agriculture, Ecosystems & Environment
638 114, 296–310. https://doi.org/10.1016/j.agee.2005.11.017

639 O'Hara, R.B., Sillanpää, M.J., 2009. A review of Bayesian variable selection methods: what,
640 how and which. Bayesian Analysis 4, 85–118. https://doi.org/10.1214/09-BA403

641 Parker, S.S., Cohen, B.S., Moore, J., 2018. Impact of solar and wind development on
642 conservation values in the Mojave Desert. PLOS ONE 13, e0207678.
643 https://doi.org/10.1371/JOURNAL.PONE.0207678

644 Pennsylvania General Assemby, 2004. Public Utilities. 66S814.

645 Perry, M.T., 2013. rasterstats.

646 Pisinaras, V., Wei, Y., Bärring, L., Gemitzi, A., 2014. Conceptualizing and assessing the effects
647 of installation and operation of photovoltaic power plants on major hydrologic budget

constituents. Science of The Total Environment 493, 239–250. https://doi.org/10.1016/j.scitotenv.2014.05.132

Plummer, M., 2017. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Queiroz, C., Beilin, R., Folke, C., Lindborg, R., 2014. Farmland abandonment: threat or opportunity for biodiversity conservation? A global review. Frontiers in Ecology and the Environment 12, 288–296. https://doi.org/10.1890/120348

R Core Team, 2021. R: A language and environment for statistical computing.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., Vilariño, M.V., 2018. Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development., in: Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M.I., Lonnoy, E., T. Maycock, T., M. Tignor, M., T. Waterfield, T. (Eds.), Global Warming of 1.5°C. An IPCC Special Report. p. 82pp.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation, in: ArXiv. pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sacchelli, S., Garegnani, G., Geri, F., Grilli, G., Paletto, A., Zambelli, P., Ciolli, M., Vettorato, D., 2016. Trade-off between photovoltaic systems installation and agricultural practices on arable lands: An environmental and socio-economic impact analysis for Italy. Land Use Policy 56, 90-99. https://doi.org/10.1016/j.landusepol.2016.04.024

Sinha, P., Hoffman, B., Sakers, J., Althouse, L., 2018. Best practices in responsible land use for improving biodiversity at a utility-scale solar facility. Case Studies in the Environment 2, 1–12. https://doi.org/10.1525/cse.2018.001123

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 583–639. https://doi.org/10.1111/1467-9868.00353

Tenan, S., O'Hara, R.B., Hendriks, I., Tavecchia, G., 2014. Bayesian model selection: The steepest mountain to climb. Ecological Modelling 283, 62–69. https://doi.org/10.1016/j.ecolmodel.2014.03.017

Trisos, C.H., Merow, C., Pigot, A.L., 2020. The projected timing of abrupt ecological disruption from climate change. Nature 580, 496–501. https://doi.org/10.1038/s41586-020-2189-9

Tsoutsos, T., Frantzeskaki, N., Gekas, V., 2005. Environmental impacts from the solar energy technologies. Energy Policy 33, 289–296. https://doi.org/10.1016/S0301-4215(03)00241-6

United States Government, 2021. The United States of America Nationally Determined Contribution Reducing Greenhouse Gases in the United States: A 2030 Emissions Target.

U.S. Energy Information Administration, 2022. Annual energy outlook 2022. Washington, D.C.

Virginia General Assembly, 2020. Virginia Electric Utility Regulation Act.

Visser, E., Perold, V., Ralston-Paton, S., Cardenal, A.C., Ryan, P.G., 2019. Assessing the impacts of a utility-scale photovoltaic solar energy facility on birds in the Northern Cape, South Africa. Renewable Energy 133, 1285–1294. https://doi.org/10.1016/J.RENENE.2018.08.106

West Virginia Legislature, 2015. An act to repeal S24-2F-1 – S24-2F-12 of the Code of West Virginia, 1931, as amended, all relating to repealing certain provisions of the Alternative and Renewable Energy Portfolio Act. H.B. 2001.

Wickham, J., Stehman, S. v., Gass, L., Dewitz, J.A., Sorenson, D.G., Granneman, B.J., Poss, R. v., Baer, L.A., 2017. Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). Remote Sensing of Environment 191, 328–341. https://doi.org/10.1016/J.RSE.2016.12.026

Wu, G., Leslie, E., Sawyerr, O., Cameron, D., Brand, E., Cohen, B., Allen, D., Ochoa, M., Olson, A., 2020. Environmental Research Letters 15(7), 074044. https://doi.org/10.1088/1748-9326/ab87d1

7. Tables

Table 1. Landscape variables hypothesized to be associated with the probability of solar array construction used in predictive models. For each variable we provide a description of the variable, the format from which the data was sampled, and the original source of the data. For raster data, the mean value within polygons was extracted. For vector data, the value of the feature within which arrays occurred was used.

Variable	Description	Format	Source
Slope	Mean slope derived from 10 m digital elevation model	Raster	3DEP National Map (USGS, 2022)
Year	Year of Sentinel-2 image in which array was first detected	Scalar	Authors
Road Distance	Distance (m) to nearest local (S1400) or secondary (S1200) road	Vector	USA Roads (USCB, 2021)
Line Distance	Distance (m) to nearest electric power transmission line	Vector	U.S. Electric Power Transmission Lines (DHS, 2022)
Gap Status	GAP protected area status code (1 = High, 5 = Low)	Vector	USGS Protected Area Dataset of the U.S.
Housing	2010 Housing density (km^{-2}) of the census tract	Vector	U.S. Census Bureau
Income	2010 median household income of the census tract	Vector	U.S. Census Bureau
Population	2010 population of census tract	Vector	U.S. Census Bureau
Cultivated	Percentage of pixels identified as cropland in 2016	Raster	Cropland Data Layer (USDA-NASS 2021)
Tree Cover	Percentage of pixels identified as tree canopy in 2016	Raster	National Land Cover Database (Dewitz & USGS, 2021)
Impervious	Percentage of pixels identified as impervious surface in 2016	Raster	National Land Cover Database (Dewitz & USGS, 2021)
Open	Percentage of pixels identified as grassland, shrub, open-developed, or low-intensity developed (i.e., lawns) in 2016	Raster	National Land Cover Database (Dewitz & USGS, 2021)
Farm Score*	Agricultural soil suitability score (1 = High, 4 = Low)	Raster	SSURGO Farmland Class (NRCS, 2021)
Latitude	Latitude (m) of array centroid	Scalar	

* We reclassified the Prime Farmland data from the Soil Survey Geographic Database (NRCS, 2021) into four classes to represent the potential suitability of soils to cultivation.

712 **Table 2**. Characteristics of solar arrays within each state. Table displays the total area and
713 proportion of each state occupied by solar arrays, mean array size, 95% credible interval around
714 rate of increase (Δ) and biodiversity selection (β) coefficeints. Bold text indicates credible
715 intervals that did not overlap zero.

State	*Solar area* (km^2)	Size ± σ (km^2)	Δ 95% CI (km^2/yr)	β 95% CI
DE	9.01 (1.79e-4%)	0.291 ± 0.52	**7.29e-4 – 2.08e-3**	-1.25e-3 – 4.66e-3
MD	89.05 (3.54e-4%)	0.425 ± 0.68	**4.35e-3 – 5.68e-3**	-1.25e-3 – 2.47e-3
NY	99.68 (0.82e-4%)	0.376 ± 0.46	**3.91e-4 – 2.26e-3**	-6.54e-4 – 2.32e-3
PA	37.45 (0.32e-4%)	0.395 ± 0.67	-7.05e-5 – 1.30e-3	-4.57e-4 – 3.85e-3
VA	274.17 (2.69e-4%)	0.862 ± 1.33	**5.57e-3 – 6.95e-3**	-8.06e-4 – 2.06e-3

716

717 **Table 3**. Posterior distributions for coefficients and model inclusion probability for all covariates
718 considered as predictors of the rate parameter for Weibull distribution of time to solar
719 development, as estimated by hierarchical Bayesian models. Bold text indicates covariates with
720 strong evidence supporting inclusion in the model.

Coefficient	$\bar{\beta}_{jfull} \pm \sigma$	$P(w_j)$	$\bar{\beta}_j \pm \sigma$	$P(\beta_j = 0)$
Impervious	-11.5 ± 0.861	0.00	-	-
Open	-0.039 ± 0.164	0.191	0.011 ±0.375	0.479
Tree Cover	-1.53 ± 0.13	0.667	-1.48 ±0.155	0.00
Cultivated	0.239 ± 0.123	1.00	0.925 ±0.246	0.00
Farm Score	1.10 ± 0.136	0.667	1.41 ±0.208	0.00
√Slope	-2.96 ± 0.303	0.333	-2.95 ±0.283	0.00
log(GAP Status)	-1.30 ± 0.402	1.00	-1.51 ±0.355	0.00
log(Line Distance)	-1.79 ± 0.151	0.00	-	-
log(Road Distance)	0.348 ± 0.138	1.00	0.919 ±0.188	0.00
log(Population)	0.293 ± 0.17	0.035	0.054 ±0.15	0.629
Income	-0.136 ± 0.273	0.036	0.031 ±0.189	0.472
Latitude	-2.32 ± 0.408	1.00	-2.24 ±0.565	0.00

721

8. Figure Legends

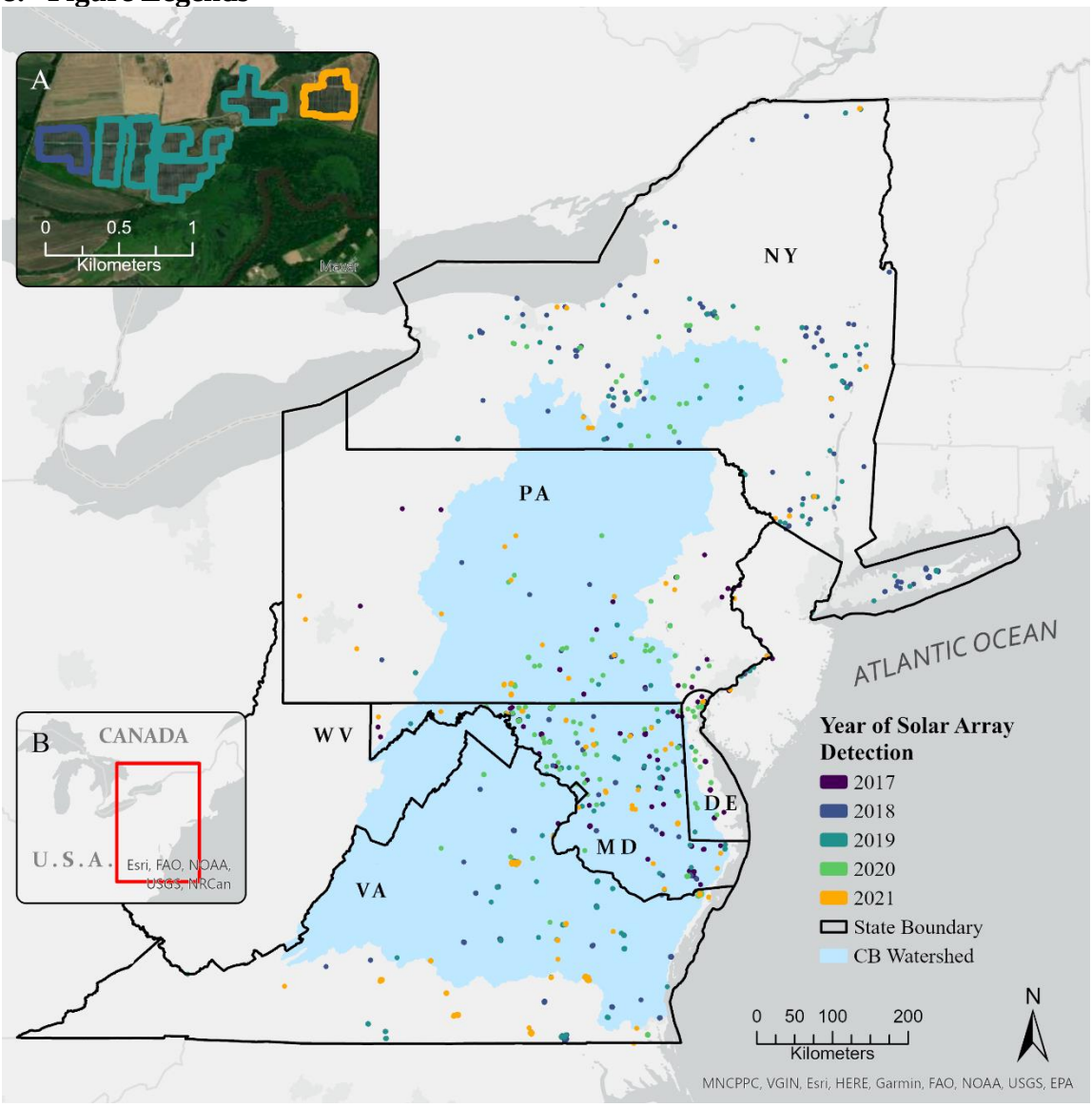
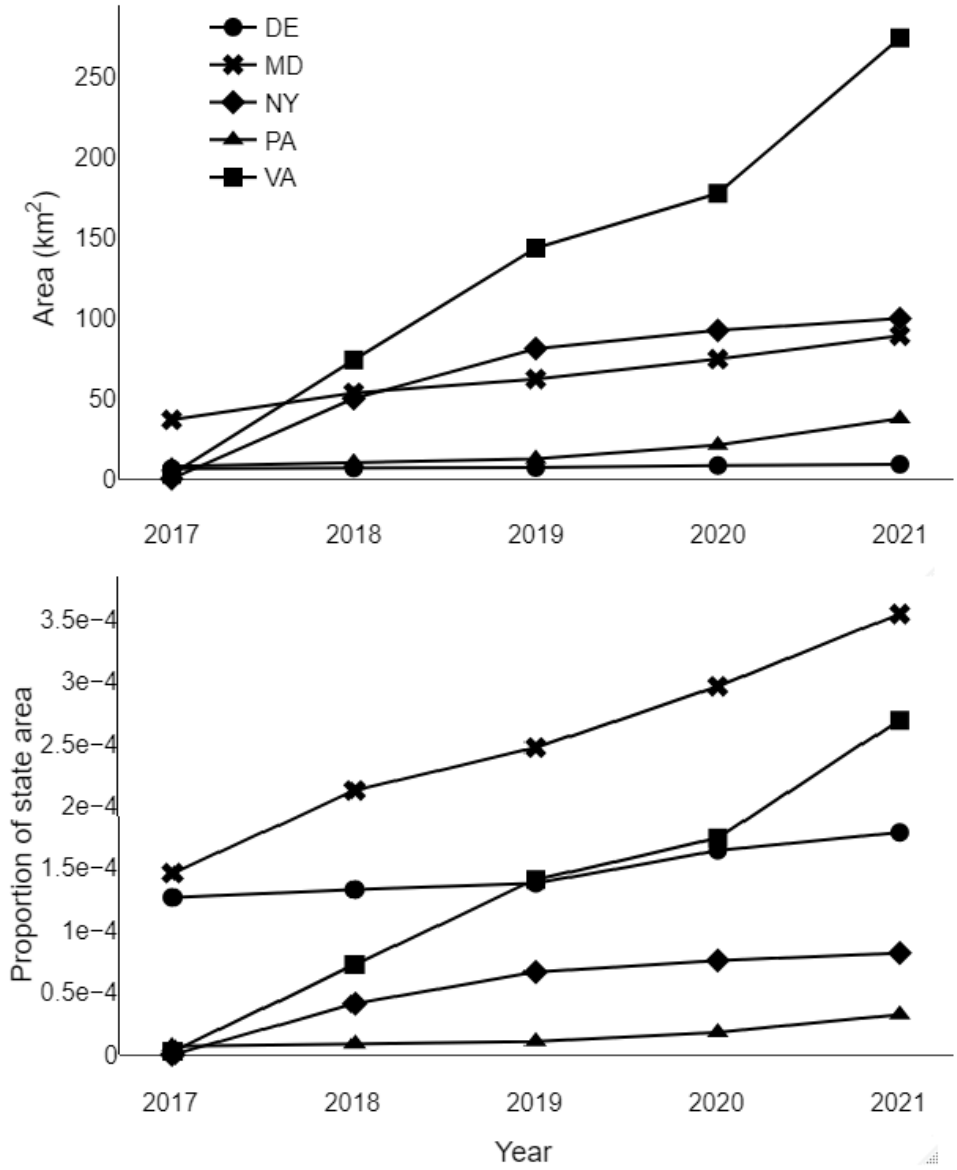


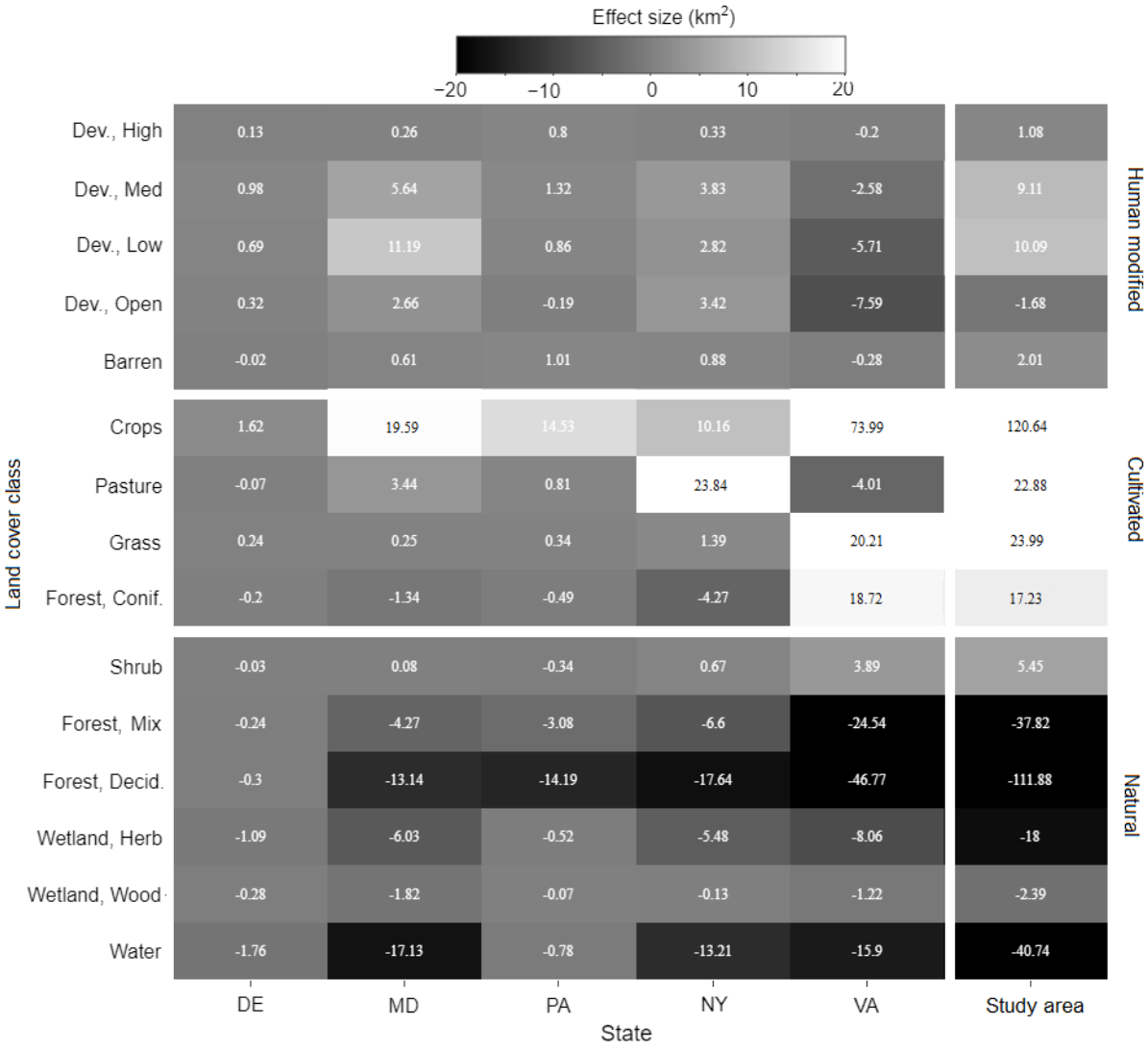
Figure 1. Polygons representing ground-mounted photovoltaic arrays present in 2021 mapped by a U-Net image segmentation model using Sentinel-2 satellite data. Arrays were mapped in each of six states overlapping the Chesapeake Bay watershed (blue). Colors indicate the year in which arrays were constructed. Map insets show (A) the detailed solar array polygons generated by the U-Net model, and (B) the location of the study area in the United States.



730

731 **Figure 2**. Solar buildout within the study area over time. Plots show (a) the total area of solar
732 arrays within each state, and (b) the area of solar arrays proportional to the total area of each
733 state from 2017 – 2021.

734



735

736 **Figure 3**. Differences between the observed and expected area (km^2) of 15 different NLCD
737 classes converted to solar arrays within the entire study area, and five component states. Classes
738 are grouped according to type of human modification: Natural; Cultivated; and Human modified.
739 Expected values were calculated based on observed proportions of each class. Positive values
740 (white) indicate more area converted to solar arrays than expected, negative (black) values
741 indicate less area converted to solar arrays than expected.

742

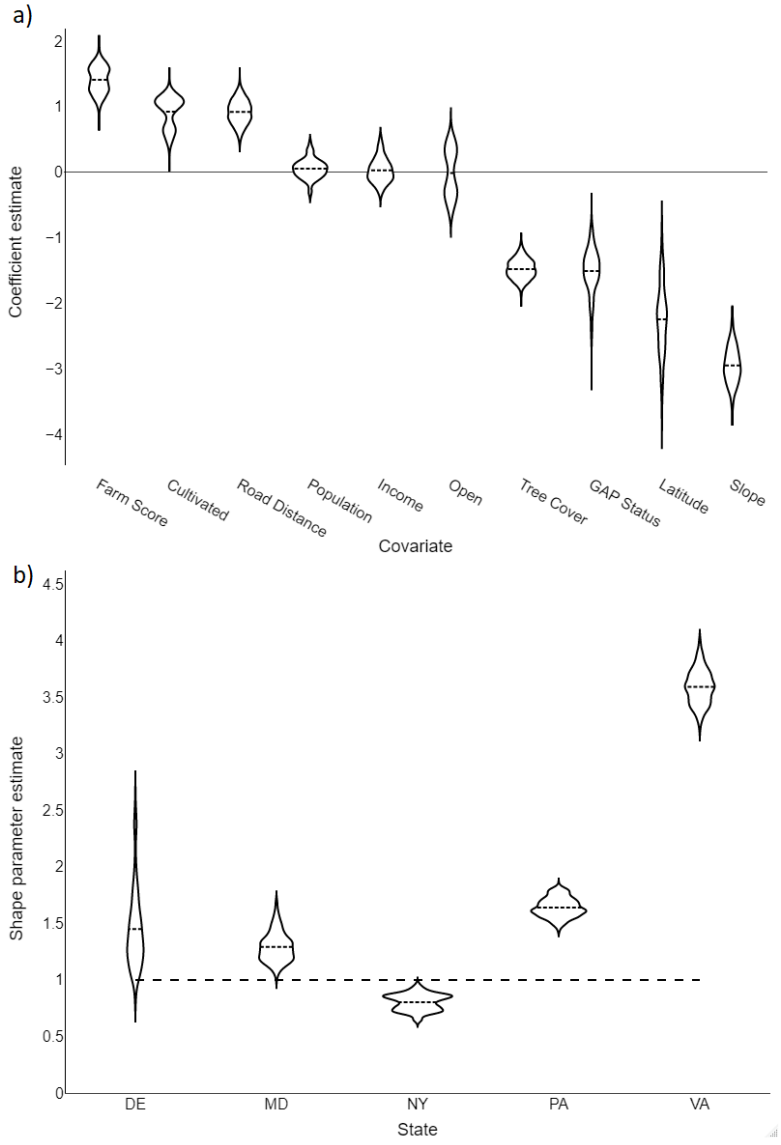
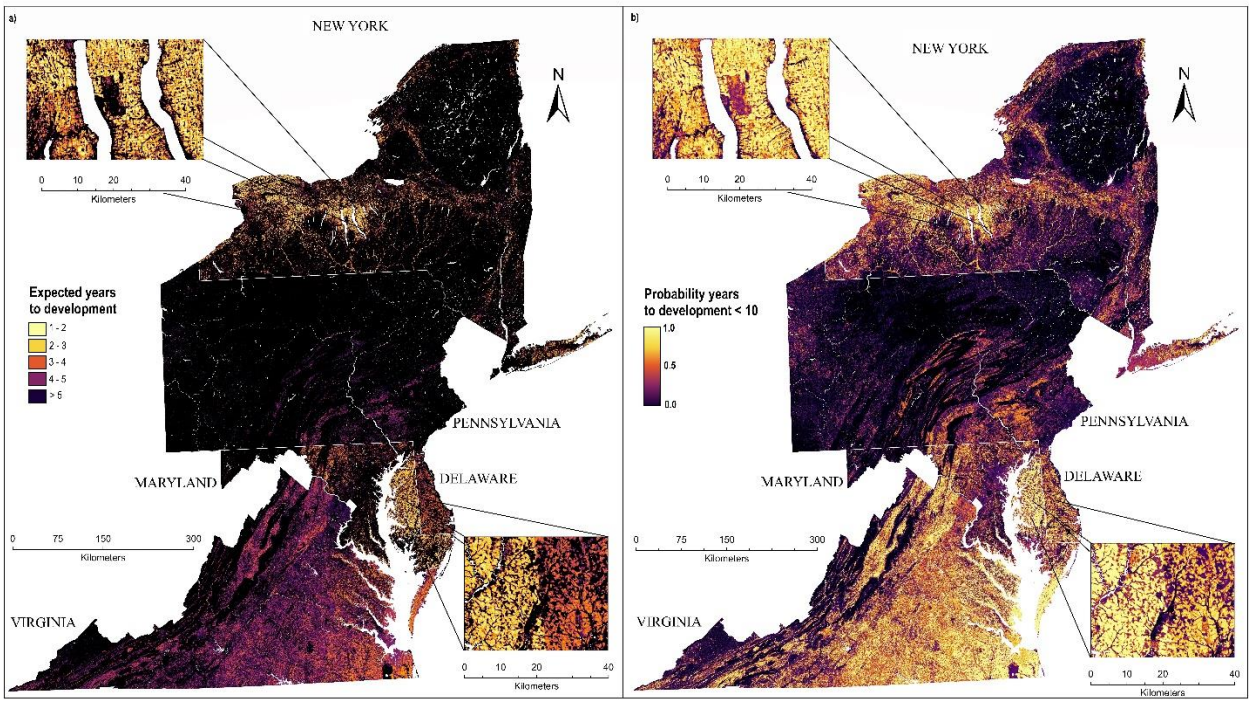


Figure 4. Posterior parameter estimates for a Weibull model of time to solar construction data within the study area. Plots show (a) covariate coefficients of the Weibull rate parameter in linear models of Weibull rate parameter and (b) per-state Weibull shape parameters. A shape parameter values above and below 1 (dotted line) indicate accelerating and decelerating rates of solar buildout over time, respectively.



751

752 **Figure 5**. Expected time to solar development across the study area. Maps show, at 30 m
753 resolution, a) the expected time to solar development and b) the cumulative probability of
754 development after 10 years, based on landcover covariates and mean parameter estimates from
755 the fitted joint binomial-Weibull model. State boundaries are shown as white lines. Insets
756 illustrate spatial variation in the relative development probability at fine scale within states.

757