

1 **Supporting Information.** Elahi, R., Edmunds, P.J., Gates, R.D., Kuffner, I.B., Barnes, B.B.,  
2 Chollett, I., Courtney, T.A., Guest, J.R., Lenz, E.A., Toth, L.T., Viehman, T.S., Williams, I.D.  
3 2021. Coral reef oases are consistently associated with high light attenuation. Ecological  
4 Applications.

5 **Appendix S2.**

6 Any use of trade, firm, or product names is for descriptive purposes only and does not  
7 imply endorsement by the U.S. Government.

## 8 Section S1. Statistical Model

9 We were interested in understanding the probability of occurrence of coral reef oases, and  
 10 in particular, how environmental covariates mediated their probability of occurrence. As  
 11 stated in the *Methods*, we defined an oasis as a reef site that exhibited higher coral cover  
 12 relative to reef sites within a defined area. These areas were divided into 2.5 arcminute grid  
 13 cells ( $\sim 21.2 \text{ km}^2$ ), and individual sites ( $< 100 \text{ m}^2$ ) were sampled within these grid cells. The  
 14 probability of occurrence can be separated into the true probability of occurrence of an oasis  
 15 ( $\psi$ ), and the probability of detecting an oasis ( $p$ ). The true probability of occurrence ( $\psi$ )  
 16 was modeled using a deterministic equation with the pertinent environmental covariates,  
 17 and the detection probability ( $p$ ) was modeled using a binomial distribution with the  
 18 number of trials (sampled reefs) and successes (oases). Our modeling approach is based  
 19 on species occupancy models (Mackenzie et al. 2002), which estimate species occupancy  
 20 (i.e., occurrence) when detection probabilities are less than one.

21 Our data set was the number of oases ( $y_{ij}$ ) observed in a given cell  $i$  nested within sub-  
 22 region  $j$ , given  $n_{ij}$  sampling occasions. We wished to predict the true probability of  
 23 occurrence of an oasis for cell  $i$ . We defined the unobserved, true state of cell  $i$  as  $z_{ij} = 1$  if it  
 24 had an oasis, and  $z_{ij} = 0$  if it did not. Then we modeled the data  $y_{ij}$ , the number of times  
 25 we observed an oasis given  $n_{ij}$  sampling occasions as:

$$y_{ij} \sim \begin{cases} 0, & \text{if } z_{ij} = 0 \\ \text{Binomial}(n_{ij}, p_j), & \text{if } z_{ij} = 1 \end{cases}$$

26 which states that we will never detect an oasis site if the cell does not have one, but if the  
 27 cell does have an oasis, we will detect it with probability  $p_j$ , estimated as a group-level  
 28 intercept for sub-region  $j$ , designated as  $h(\eta_j)$  below.

29 Next, we modeled the process governing the true state  $z_{ij}$ :

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij})$$

30 We used a Bernoulli distribution because the random variable  $z_{ij}$  can take on values of 0 or  
 31 1. The frequency of these values is determined by the true probability of occurrence,  $\psi_{ij}$ .  
 32 We modeled  $\psi_{ij}$  using a deterministic model,  $g(\alpha_j, \boldsymbol{\beta}, \mathbf{x}_{ij})$ , where  $\alpha_j$  represented an intercept  
 33 for sub-region  $j$ ,  $\boldsymbol{\beta}$  represented a vector of coefficients, and  $\mathbf{x}_{ij}$  represented a vector of  
 34 the measured covariates for cell <sub>$ij$</sub> . The covariates were assumed to be measured without  
 35 error and thus were not treated as random variables in our model. We used an inverse  
 36 logit function because it returns continuous values from 0 to 1. Finally, we calculated the  
 37 posterior probability of our random variables conditional on our data using the following  
 38 Bayesian hierarchical model:

$$\begin{aligned}
 [\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \mu_\alpha, \sigma_\alpha, \mu_\eta, \sigma_\eta | \mathbf{n}, \mathbf{y}] &\propto \prod_{i=1}^{890} \prod_{j=1}^{32} [y_{ij} | n_{ij}, h(\eta_{ij}) z_{ij}] && \text{(detection model)} \\
 &\times [z_{ij} | g(\alpha_j, \boldsymbol{\beta}, \mathbf{x}_{ij})] && \text{(occurrence model)} \\
 &\times [\alpha_j | \mu_\alpha, \sigma_\alpha] && \text{(occurrence hyperprior)} \\
 &\times [\eta_j | \mu_\eta, \sigma_\eta] && \text{(detection hyperprior)} \\
 &\times [\boldsymbol{\beta}] [\mu_\alpha] [\sigma_\alpha] [\mu_\eta] [\sigma_\eta] && \text{(priors)} \\
 g(\alpha_j, \boldsymbol{\beta}, \mathbf{x}_{ij}) &= \text{invlogit}(\alpha_j + \boldsymbol{\beta} \mathbf{x}_{ij}) \\
 h(\eta_{ij}) &= \text{invlogit}(\eta_j)
 \end{aligned}$$

39 with the following priors:

$$\beta \sim \text{Normal}(0, 1)$$

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$$

$$\mu_\alpha \sim \text{Normal}(-1, 1)$$

$$\sigma_\alpha \sim \text{Exponential}(1)$$

$$\eta_j \sim \text{Normal}(\mu_\eta, \sigma_\eta)$$

$$\mu_\eta \sim \text{Normal}(-1, 1)$$

$$\sigma_\eta \sim \text{Exponential}(1)$$

40 We used weakly regularizing priors for the slope coefficients ( $\beta$ ), noting that the envi-  
 41 ronmental covariates were standardized to have a mean of 0 and standard deviation of 1  
 42 (Fig. S1A). We also chose weakly regularizing priors for the hyperpriors  $\mu_\alpha$  (Fig. S1B) and  
 43  $\sigma_\alpha$  (Fig. S1C) so that their resulting group-level intercepts ( $\alpha_j$ ) for the true probability of  
 44 occurrence peaked between 0 and 0.3, and then declined steadily towards 1 (Fig. S2A). We  
 45 chose to put more weight on probabilities less than 0.5 because oases are, by definition, rare  
 46 occurrences. The same priors were used for hyperpriors  $\mu_\eta$  and  $\sigma_\eta$ , for the same reasons.  
 47 We visualized our prior predictive distributions to ensure that the resulting relationships  
 48 between the true probability of occurrence and a standardized coefficient were reasonable  
 49 (Fig. S2B).

Section S2. Supporting Figures

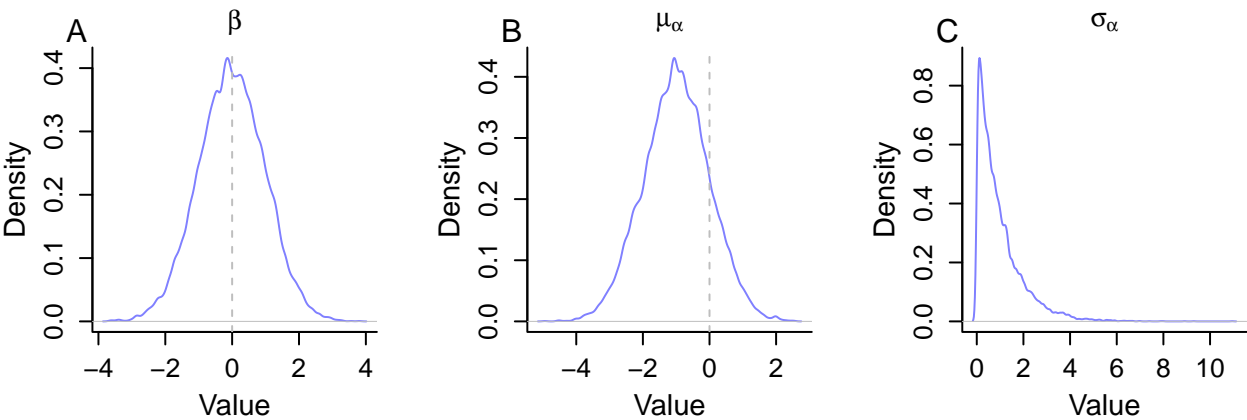
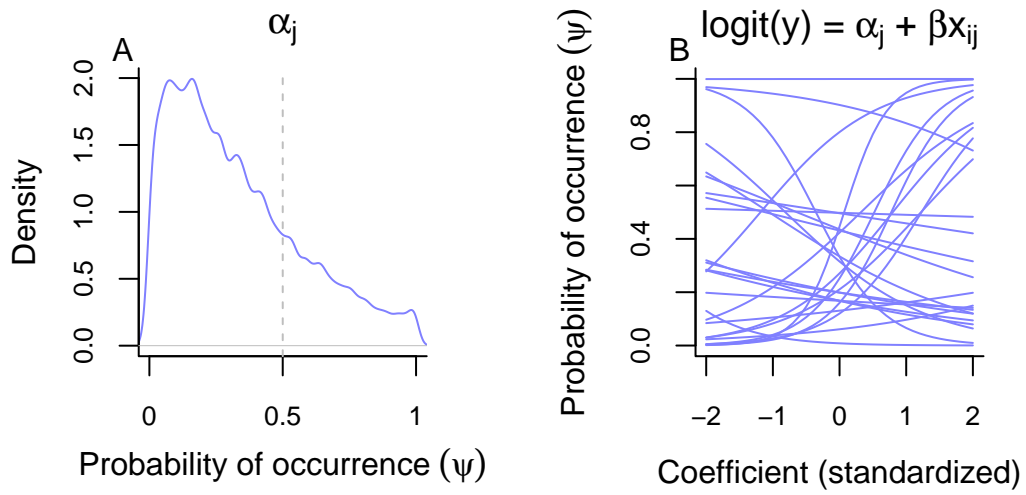


Figure S1. Predictive distributions based on 10000 simulated draws for the priors from the hierarchical model. Parameters are on the logit scale.



55

56 Figure S2. Predictive distributions based on 10000 simulated draws for the priors from the  
 57 hierarchical model. In (A), the probability of oasis occurrence is back-transformed from  
 58 the logit scale. In (B), 25 randomly selected relationships between the true probability of  
 59 occurrence and a standardized covariate derived from the prior predictive distributions.

```

model{

  # priors for occurrence model
  for(i in 1:nX){
    beta[i] ~ dnorm(0, 1)
  }

  # hyper-priors for occurrence model
  mu.alpha ~ dnorm(-1, 1)
  sigma.alpha ~ dexp(1)
  tau.alpha <- 1/sigma.alpha^2
  for(j in 1:y.n.sites){
    alpha[j] ~ dnorm(mu.alpha, tau.alpha)
  }

  # hyper-priors for detection model
  mu.det ~ dnorm(-1, 1)
  sigma.det ~ dexp(1)
  tau.det <- 1 / sigma.det^2
  for(j in 1:y.n.sites){
    eta[j] ~ dnorm(mu.det, sigma.det)
  }

  # likelihood
  for(i in 1:N){

```

```

# occurrence model

logit(psi[i]) <- alpha[y.group[i]] + inprod(beta[], X[i, ] )
z[i] ~ dbern(psi[i])

# detection model

logit(p[i]) <- eta[y.group[i]]
mu.p[i] <- z[i] * p[i]
y[i] ~ dbin(mu.p[i], n[i])

# simulate new data, conditional on model parameters
y.new[i] ~ dbin(mu.p[i], n[i])

# pearson chi-square discrepancy for a binomial
# e is small value to avoid division by zero
chi2b.data[i] <- ((y[i] - mu.p[i] * n[i]) /
  sqrt((mu.p[i] + e) * n[i] * (1 - mu.p[i] - e)))^2
chi2b.sim[i] <- ((y.new[i] - mu.p[i] * n[i]) /
  sqrt((mu.p[i] + e) * n[i] * (1 - mu.p[i] - e)))^2

# freeman-tukey discrepancy for a binomial
ftd.data[i] <- (sqrt(y[i]) - sqrt(p[i] * z[i] * n[i]))^2
ftd.sim[i] <- (sqrt(y.new[i]) - sqrt(p[i] * z[i] * n[i]))^2

}

# bayesian p-value for chi-square discrepancy

```



```
d.chi2b.data <- sum(chi2b.data)
d.chi2b.sim <- sum(chi2b.sim)
p.chi2b <- step(d.chi2b.sim - d.chi2b.data)

# bayesian p-value for freeman-tukey discrepancy
d.ftd.data <- sum(ftd.data)
d.ftd.sim <- sum(ftd.sim)
p.ftd <- step(d.ftd.sim - d.ftd.data)

}
```

## 61 **References**

- 62 MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm,  
63 C. A. (2002). Estimating site occupancy rates when detection probabilities are less than  
64 one. *Ecology*, 83(8), 2248-2255.