Supplementary material

Section S1 provides necessary assumptions to establish Theorems 1 and 2. Sections S2 and S3 provide the proofs of Theorems 1, 2, and 3. Section S4 describes the proposed cluster-weighted bootstrap for the average treatment effect on the treated (ATT). Section S5 extends the matching framework to unit-level treatment assignments. Section S6 presents additional figures from the analysis of conservation policy effects on marine biodiversity.

S1 Additional assumptions for Theorems 1 and 2

Assumption S1. (Conditions for $\hat{\tau}$)

(i) $\{Y_{ir}\}_{i=1,r=1}^{N,R}$ follows Model (1) that are independent between clusters but dependent within cluster.

(ii) S is continuously distributed on a compact and convex support $\mathbb{S}$. The density of S is bounded and bounded away from zero on $\mathbb{S}$.

(iii) A is independent of $(Y(0), Y(1))$ conditional on $S = s$ for almost every s. There exists a positive constant c such that $Pr(A = 1|S = s) \in (c, 1 - c)$ for almost every s.

(iv) For each $a \in \{0,1\}$, $\mu(a, s)$ and $\sigma^2(a, s)$ are Lipschitz in $\mathbb{S}$, $\sigma^2(a, s)$ is bounded away from zero on $\mathbb{S}$ and $E[Y^4|A = a, S = s]$ is bounded uniformly on $\mathbb{S}$.

(v) $E(K_M(i, j)^q)$ is uniformly bounded over N.

The bias-corrected estimators involve the estimation of conditional bias terms B_M and B_M^t. Following Abadie and Imbens [1], we assume the required condition for estimation on outcome mean functions in clustered data.

Assumption S2. (Conditions for $\mu(s, a)$) Let $\lambda = (\lambda_1, \dots, \lambda_k)'$ be a k-dimensional vector of non-negative integers, $\partial^{\lambda} a(s) = \partial^{\sum_{l=1}^k \lambda_l} a(s)/\partial s_1^{\lambda_1} \dots \partial s_k^{\lambda_k}$ and $|a(\cdot)|_m = \max_{\sum_{l=1}^k \lambda_l \le m} \sup_{s \in \mathbb{S}} |\partial^{\lambda} a(s)|$. For each $a \in \{0,1\}$ and λ satisfying $\sum_{l=1}^k \lambda_l = k$, the derivative $\partial^{\lambda}\mu(s, a)$ exists and satisfies $\sup_{s \in \mathbb{X}} |\partial^{\lambda}(\mu(s, a))| \le C$ for some $C > 0$. Furthermore, $\hat{\mu}(s, a)$ satisfies $|\hat{\mu}(\cdot, a) - \mu(\cdot, a)|_{m-1} = o_p(N^{-1/2+1/k})$ for each $a \in \{0,1\}$.

This assumption S2 guarantees a fast convergence rate on $\hat{B}_M$ and thus contributes to establishing the remarkable result with certain regularity conditions required in Abadie and Imbens [1],

$$\sqrt{N}(\hat{B}_M - B_M) \overset{p}{\to} 0. \tag{S1}$$

Assumption S2 is also necessary for the bias-corrected ATT estimator $\hat{\tau}^t$ to ensure $\sqrt{N}(\hat{B}_M^t - B_M^t) \overset{p}{\to} 0$ under certain conditions.

S2 Proofs of Theorem 1 and 2

Let S be a generic variable for matching which could be at cluster-level or unit-level. The original estimator of average treatment effect (ATE) is

$$\hat{\tau}_{\mathrm{mat}} = N^{-1}\sum_{i=1}^{N}\sum_{r=1}^{R}(2A_{ir}-1)\left\{1+\frac{1}{M}K_M(i,r)\right\}Y_{ir}.$$

According to [2], we write $\sqrt{N}(\hat{\tau}_{\mathrm{mat}} - B_M - \tau) = \sqrt{N}(\overline{\tau(S)} - \tau) + \sqrt{N}E_M$, where

$$\overline{\tau(S)} - \tau = \frac{1}{N}\sum_{i=1}^{N}\sum_{r=1}^{R}\mu_1(S_{ir}) - \mu_0(S_{ir}) - \tau,$$
$$E_M = \frac{1}{N}\sum_{i=1}^{N}E_{M,i} = \frac{1}{N}\sum_{i=1}^{N}\sum_{r=1}^{R}(2A_r-1)\left\{1+\frac{K_M(i,r)}{M}\right\}\{Y_{ir} - \mu_{A_r}(S_{ir})\}.$$

Then for the first part $\sqrt{n}(\overline{\tau(S)} - \tau)$, by a standard central limit theorem,

$$\sqrt{N}(\overline{\tau(S)} - \tau) \xrightarrow{d} N(0, V^{\tau(S)}).$$

Consider the distribution of $\sqrt{N}E_M/\sqrt{V^E}$, we adopt the results shown in [2] that the moments of $K_M(i,r)$ are bounded uniformly in N. For a given S, A, the Lindeberg-Feller condition requires that

$$\frac{1}{NV^E}\sum_{i=1}^{N}\mathbb{E}[(E_{M,i}^2)\mathbb{1}\{|E_{M,i}| \geq \eta\sqrt{NV^E}\}|S,A] \to 0$$

for all $\eta > 0$. By the same routine, this condition holds by using Hölder's and Markov's inequalities,

$$\begin{aligned}
&\mathbb{E}[(E_{M,i}^2)\mathbb{1}\{|E_{M,i}| \geq \eta\sqrt{NV^E}\}|X,A]\\
&\leq \mathbb{E}[(E_{M,i}^4|S,A))^{1/2}(\mathbb{E}[\mathbb{1}\{|E_{M,i}| \geq \eta\sqrt{NV^E}\}|S,A])^{1/2}\\
&\leq \mathbb{E}[(E_{M,i}^4|S,A))^{1/2}(\Pr(|E_{M,i}| \geq \eta\sqrt{NV^E}|S,A))\\
&\leq \mathbb{E}[(E_{M,i}^4|S,A))^{1/2}\frac{\mathbb{E}[(E_{M,i})^2|S,A]}{\eta^2 NV^E}.
\end{aligned}$$

Let $\bar{\sigma}^2 = \sup_{a,s}\sigma^2(s,a) < \infty$, $\underline{\sigma}^2 = \inf_{a,s}\sigma^2(s,a) > 0$ and $\bar{C} = \sup_{a,s}\mathbb{E}\left[\{Y_{ir} - \mu_{A_r}(S_{ir})\}^4|S_{ir} = s, A_r = a\right] < \infty$, then the condition is bounded as

$$\frac{1}{NV^E}\sum_{i=1}^{N}\mathbb{E}[(E_{M,i}^2)\mathbb{1}\{|E_{M,i}| \geq \eta\sqrt{NV^E}\}|S,A] \leq \frac{\bar{\sigma}^2\bar{C}^{1/2}}{\eta^2\underline{\sigma}^4 N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{r=1}^{R}(1+M^{-1}K_M(i,r))\right)^4\right).$$

By central limit theorem for dependent variables [3], with finite V^E, we have

$$\sqrt{N}E_M \xrightarrow{d} N(0, V^E).$$

Therefore, the matching estimator $\hat{\tau}_{\mathrm{mat}}$ follows asymptotic normal distribution with finite variance $(V^E + V^{\tau(S)})/N$, after ignoring the conditional bias term.

For the estimator of average treatment effect on the treated (ATT),

$$\hat{\tau}^t_{\mathrm{mat}} = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}\left\{A_r - (1-A_r)\frac{K_M(i,r)}{M}\right\}Y_{ir},$$

where N_1 is the number of units in treatment group. After ignoring the conditional bias term B^t_M,

$$\begin{aligned}
\tilde{\tau}^t &= \overline{\tau(X)}^t - B^t_M\\
&= \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}A_{ir}\{\mu(S_{ir},1) - \mu_0(S_{ir})\} + \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}\left\{A_r - (1-A_r)\frac{K_M(i,r)}{M}\right\}\{Y_{ir} - \mu_{A_r}(S_{ir})\}
\end{aligned}$$

$$= \frac{1}{N_1}\left[\sum_{r=1}^{R} A_r\{\mu(S_{ir},1) - \mu_0(S_{ir})\} + \left\{A_r - (1-A_r)\frac{K_M(i,r)}{M}\right\}\{Y_{ir} - \mu_{A_r}(S_{ir})\}\right]$$
$$= \overline{\tau(S)}^t + E_M^t,$$

where

$$\overline{\tau(S)}^t = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R} A_r\{\mu(S_{ir},1) - \mu_0(S_{ir})\},$$
$$E_M^t = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}\left\{A_r - (1-A_r)\frac{K_M(i,r)}{M}\right\}\{Y_{ir} - \mu_{A_r}(S_{ir})\}.$$

Consider $\sqrt{N_1}(\hat{\tau}^t_{\text{mat}} - B^t_M - \tau^t) = \sqrt{N_1}(\overline{\tau(S)}^t - \tau^t) + \sqrt{N_1}E^t_M$, according to the similar procedure described above, the first part $\sqrt{N_1}(\overline{\tau(S)}^t - \tau^t) \overset{d}{\to} N(0, V^{\tau(S),t})$ by the standard central limit theorem. The second part includes

$$E_M^t = \sum_{r=1}^{R}\left\{A_r - (1-A_r)\frac{K_M(i,r)}{M}\right\}\{Y_{ir} - \mu_{A_r}(S_{ir})\}.$$

By the central limit theorem for dependent variables, we have

$$\sqrt{N}E_M^t \overset{d}{\to} N(0, V^{E,t}).$$

The matching estimator $\hat{\tau}^t_{\text{mat}}$ thus follows an asymptotic normal distribution with finite variance $(V^{E,t} + V^{\tau(S),t})/N_1$, after ignoring the conditional bias term.

Under certain regularity conditions in [1] and the condition for outcome mean functions [4], we have $\sqrt{N}(\hat{B}_M - B_M) \overset{p}{\to} 0$ and $\sqrt{N}(\hat{B}^t_M - B^t_M) \overset{p}{\to} 0$. Therefore, the above asymptotic normality and the fast convergence of estimated condition bias terms help complete the proofs for Theorem 1 and Theorem 2.

S3 Proof of Theorem 3

For simplicity of notation, we focus on the case with balanced cluster size and reconstruct the matching estimator for convenience. Denote $N = Rn$ as the number of observations, R as the number of balanced clusters, n as the cluster size for the each cluster, and S as a unified variable that represents either the cluster-level covariates or stacked covariates, the matching estimator is

$$\hat{\tau}_{\text{mat}} = \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}Y_{rj}.$$

The matching estimator can be decomposed into three parts, $\hat{\tau}_{\text{mat}} = \overline{\tau(S)} + E_M + B_M$, where

$$\overline{\tau(X,Z)} = \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}\{\mu_1(S_{rj}) - \mu_0(S_{rj})\},$$
$$E_M = \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}(2A_r - 1)\left\{1 + \frac{K_M(r,j)}{M}\right\}\{Y_{rj} - \mu_{A_r}(S_{rj})\},$$
$$B_M = \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}(2A_r - 1)\left[\frac{1}{M}\sum_{(k,l)\in\mathcal{J}_M(r,j)}\{\mu_{1-A_r}(S_{rj}) - \mu_{1-A_r}(S_{kl})\}\right].$$

The debiased estimator for cluster weighted bootstrap is expressed as

$$
\begin{aligned}
\tilde{\tau} &= \hat{\tau}_{\text{mat}} - \hat{B}_M \\
&= \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}\{\hat{\mu}_1(S_{rj}) - \hat{\mu}_0(S_{rj})\} + \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\{Y_{rj} - \hat{\mu}_{A_r}(S_{rj})\} \\
&= \frac{1}{Rn}\sum_{r=1}^{R}\sum_{j=1}^{n}[\{\hat{\mu}_1(S_{rj}) - \hat{\mu}_0(S_{rj})\} + (2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\{Y_{rj} - \hat{\mu}_{A_r}(S_{rj})\}] \\
&= \frac{1}{R}\sum_{r=1}^{R}\tilde{\tau}_r,
\end{aligned}
$$

where $\tilde{\tau}_r = \frac{1}{n}\sum_{j=1}^{n}[\{\hat{\mu}_1(S_{rj}) - \hat{\mu}_0(S_{rj})\} + (2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\{Y_{rj} - \hat{\mu}_{A_r}(S_{rj})\}]$.

Let $\hat{e}_{rj} = Y_{rj} - \hat{\mu}_{A_r}(S_{rj})$, $\hat{\xi}_{rj} = (2A_r - 1)\{\hat{\mu}(A_r, S_{rj}) - \hat{\mu}(1 - A_r, S_{rj})\} - \tilde{\tau}$, then

$$
\tilde{\tau}_r - \tilde{\tau} = \frac{1}{n}\sum_{j=1}^{n}[(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\hat{e}_{rj} + \hat{\xi}_{rj}].
$$

To show that the cluster weighted bootstrap is a valid approach for statistical inference, we need to show the consistency of cluster weighted bootstrap variance, i.e.,

$$
E\left[\left\{\sum_{r=1}^{R} W_r^*(\tilde{\tau}_r - \tilde{\tau})|(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{S})\right\}^2\right] \overset{p}{\to} \sigma^2,
$$

where $W_r^* = M_r^*/\sqrt{R}$, and $(M_1^*, \ldots, M_R^*)$ is required to satisfy Assumption S3, e.g., it could be a vector from a multinomial distribution with equal probability. The proof follows the same strategy as Otsu and Rai [4], adjusting for the clustered case.

Assumption S3. (Conditions for W_r^*)

(i) $(W_1^*, \ldots, W_R^*)$ is exchangeable and independent of $\boldsymbol{Z} = (\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{S})$.

(ii) $\sum_{r=1}^{R}(W_r^* - \bar{W}^*)^2 \overset{p}{\to} 1$ where $\bar{W}^* = R^{-1}\sum_{r=1}^{R} W_r^*$.

(iii) $\max_{1,\ldots,R}|W_r^* - \bar{W}^*| \overset{p}{\to} 0$.

(iv) $E[W_r^{*2}] = O(R^{-1})$ for all $r = 1, \ldots, R$.

Consider $\sqrt{R}\,T^* = \sum_{r=1}^{R}(W_r^* - \bar{W}^*)(\tilde{\tau}_r - \tilde{\tau})$, it can be decomposed into three parts,

$$
\begin{aligned}
\sqrt{R}\,T^* &= \sum_{r=1}^{R}(W_r^* - \bar{W}^*)(\tilde{\tau}_r - \tilde{\tau}) \\
&= \sum_{r=1}^{R}(W_r^* - \bar{W}^*)\left[\frac{1}{n}\sum_{j=1}^{n}\{(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\hat{e}_{rj} + \hat{\xi}_{rj}\}\right] \\
&= \sqrt{R}\,(T_N^* + R_{1N}^* + R_{2N}^*),
\end{aligned}
$$

where

$$
\begin{aligned}
\sqrt{R}\,T_N^* &= \sum_{r=1}^{R}(W_r^* - \bar{W}^*)\left[\frac{1}{n}\sum_{j=1}^{n}\{(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}e_{rj} + \xi_{rj}\}\right], \\
\sqrt{R}\,R_{1N}^* &= \sum_{r=1}^{R}(W_r^* - \bar{W}^*)\left[\frac{1}{n}\sum_{j=1}^{n}(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\{\mu(A_r, S_{rj}) - \hat{\mu}(A_r, S_{rj})\}\right], \\
\sqrt{R}\,R_{2N}^* &= \sum_{r=1}^{R}(W_r^* - \bar{W}^*)\left[\frac{1}{n}\sum_{j=1}^{n}(\hat{\xi}_{rj} - \xi_{rj})\right].
\end{aligned}
$$

We need to show that $\Pr\{\sqrt{R}\,|R_{1N}^*| > \varepsilon\} \overset{p}{\to} 0$, $\Pr\{\sqrt{R}\,|R_{2N}^*| > \varepsilon\} \overset{p}{\to} 0$, and $E[(\sqrt{R}\,T_N^*)^2] \overset{p}{\to} \sigma^2$.

For $\sqrt{R}R_{1N}^*$, the Markov's inequality is leveraged to show its convergence toward 0 as $R \to \infty$.

$$
\begin{aligned}
E[(\sqrt{R}R_{1N}^*)^2] &= E\left[\left[\sum_{r=1}^{R}(W_r^* - \bar{W}^*)\gamma_r\right]^2\right] \\
&= RE[(W_1^* - \bar{W}^*)^2]\frac{1}{R}\sum_{r=1}^{R}\gamma_r^2 + R(R-1)E[(W_1^* - \bar{W}^*)(W_2^* - \bar{W}^*)] \times \frac{1}{R(R-1)}\sum_{i\neq k}\gamma_r\gamma_k
\end{aligned}
$$

where $\gamma_r = \frac{1}{n}\sum_{j=1}^{n}(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}\{\mu(A_r, S_{rj}) - \hat{\mu}(A_r, S_{rj})\}$. Since $RE[(W_1^* - \bar{W}^*)^2] = O(1), R(R-1)$ $E[(W_1^* - \bar{W}^*)(W_2^* - \bar{W}^*)] = O(1)$, $|\mu(a,\cdot) - \hat{\mu}(a,\cdot)|_{k-1} = o_p(N^{-1/2+1/k})$, and $E(K_M(i,j)^q)$ is uniformly bounded over N, we obtain $E[(\sqrt{R}R_{1N}^*)^2] \xrightarrow{p} 0$. Then the Markov inequality shows that $\Pr[\sqrt{R}|R_{1N}^*| > \varepsilon] \xrightarrow{p} 0$.

Similarly, $\Pr[\sqrt{R}|R_{2N}^*| > \varepsilon] \xrightarrow{p} 0$ by Markov inequality, since

$$
\begin{aligned}
E[(\sqrt{R}R_{2N}^*)^2] &= E\left[\left[\sum_{r=1}^{R}(W_r^* - \bar{W}^*)\left\{\frac{1}{n}\sum_{j=1}^{n}(\hat{\xi}_{rj} - \xi_{rj})\right\}\right]^2\right] \\
&= RE[(W_1^* - \bar{W}^*)^2]\frac{1}{R}\sum_{r=1}^{R}\left\{\frac{1}{n}\sum_{j=1}^{n}(\hat{\xi}_{rj} - \xi_{rj})\right\}^2 + R(R-1)E[(W_1^* - \bar{W}^*)(W_2^* - \bar{W}^*)] \\
&\quad \times \frac{1}{R(R-1)}\sum_{i\neq k}\left\{\frac{1}{n}\sum_{j=1}^{n}(\hat{\xi}_{rj} - \xi_{rj})\right\}\left\{\frac{1}{n}\sum_{l=1}^{n}(\hat{\xi}_{kl} - \xi_{kl})\right\} \\
&\xrightarrow{p} 0.
\end{aligned}
$$

For $\sqrt{R}T_N^*$, we shall show $E[(\sqrt{R}T_N^*)^2|\boldsymbol{Z}] = \sigma^2$.

$$
\begin{aligned}
E[(\sqrt{R}T_N^*)^2|\boldsymbol{Z}] &= E\left[\left[\sum_{r=1}^{R}(W_r^* - \bar{W}^*)\frac{1}{n}\sum_{j=1}^{n}\{(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}e_{rj} + \xi_{rj}\}\right]^2\right] \\
&= RE[(W_1^* - \bar{W}^*)^2]\frac{1}{R}\sum_{r=1}^{R}E\left[\left[\frac{1}{n}\sum_{j=1}^{n}\{(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}e_{rj} + \xi_{rj}\}\right]^2\right] \\
&\quad + R(R-1)\sum_{r,j\neq k,l}E[(W_1^* - \bar{W}^*)(W_2^* - \bar{W}^*)] \\
&\quad \times \frac{1}{R(R-1)}E\left[\frac{1}{n}\sum_{j=1}^{n}\{(2A_r - 1)\{1 + M^{-1}K_M(r,j)\}e_{rj} + \xi_{rj}\}\right] \\
&\quad \times E\left[\frac{1}{n}\sum_{k=1}^{n}\{(2A_k - 1)\{1 + M^{-1}K_M(k,l)\}e_{kl} + \xi_{kl}\}\right] \\
&\xrightarrow{p} \sigma^2
\end{aligned}
$$

since $RE[(W_1^* - \bar{W}^*)^2] = O(1), R(R-1)E[(W_1^* - \bar{W}^*)(W_2^* - \bar{W}^*)] = O(1)$, and σ^2 is the variance of the debiased matching estimator $\tilde{\tau}$. Therefore, it is enough to show $E\left[\left\{\sum_{r=1}^{R}W_r^*(\tilde{\tau}_r - \tilde{\tau})|\boldsymbol{Z}\right\}^2\right] \xrightarrow{p} \sigma^2$.

S4 Cluster weighted bootstrap for the average treatment effect on the treated

According to Theorem 2, under certain conditions, the bias-corrected estimator for the average treatment effect on the treated is asymptotically normal, i.e.,

$$\frac{\sqrt{N_1}(\hat{\tau}^t - \tau^t)}{\sigma^t} \xrightarrow{d} N(0,1),$$

where

$$(\sigma^t)^2 = (\sigma_1^t)^2 + (\sigma_2^t)^2,$$
$$(\sigma_1^t)^2 = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}\{A_r - (1-A_r)M^{-1}K_M(i,r)\}^2\sigma^2(X_{ir}, Z_r, A_r),$$
$$(\sigma_2^t)^2 = E[\{(\mu(X_{ir}, Z_r, 1) - \mu(X_{ir}, Z_r, 0)) - \tau^t\}^2 | A_r = 1].$$

With bias correction, we rewrite our bias-corrected estimator $\hat{\tau}^t = \hat{\tau}^t_{\text{mat}} - \hat{B}^t_M$,

$$\hat{\tau}^t = \frac{1}{N_1}\sum_{i=1}^{N}\sum_{r=1}^{R}[A_r\{Y_{ir} - \hat{\mu}(X_{ir}, Z_r, 1-A_r)\}] + (1-A_r)\frac{K_M(i,r)}{M}\{Y_{ir} - \hat{\mu}(X_{ir}, Z_r, A_r)\}$$
$$= \frac{1}{N_1}\sum_{i=1}^{N}\hat{\tau}_i^t$$

Then with suitable generated weights $\{W_i^*\}_{i=1}^N$,

$$\sqrt{N_1}(\hat{\tau}^t - \tau^t) = \sum_{i=1}^{N} W_i^*(\hat{\tau}_i^t - \hat{\tau}^t) \xrightarrow{d} N(0, (\sigma^t)^2).$$

Applying to our clustered data, we proposed the following cluster weighted bootstrap method on variance estimation for the average treatment effect on the treated.

- **Step 1**': Obtain the weighted bootstrap samples $\{\hat{\tau}_i\}_{i=1}^N$ based on the matching estimator framework.
- **Step 2**': For clustered data with N observations and R non-overlapped clusters, sample R clusters with replacement.
- **Step 3**': Include all $\{\hat{\tau}_i\}_{i=1}^N$ within selected clusters and calculate their corresponding weights $\{W_i^*\}_{i=1}^N$. One option of generating weights is to set $W_i^* = M_i^*/\sqrt{N}$, where $(M_1^*, \ldots, M_N^*)$ is a vector from a multinomial distribution with equal probability.
- **Step 4**': Obtain a bootstrap replicate as $\hat{\tau}_b^* = \sum_{i=1}^N W_i^*(\hat{\tau}_i - \hat{\tau})$.

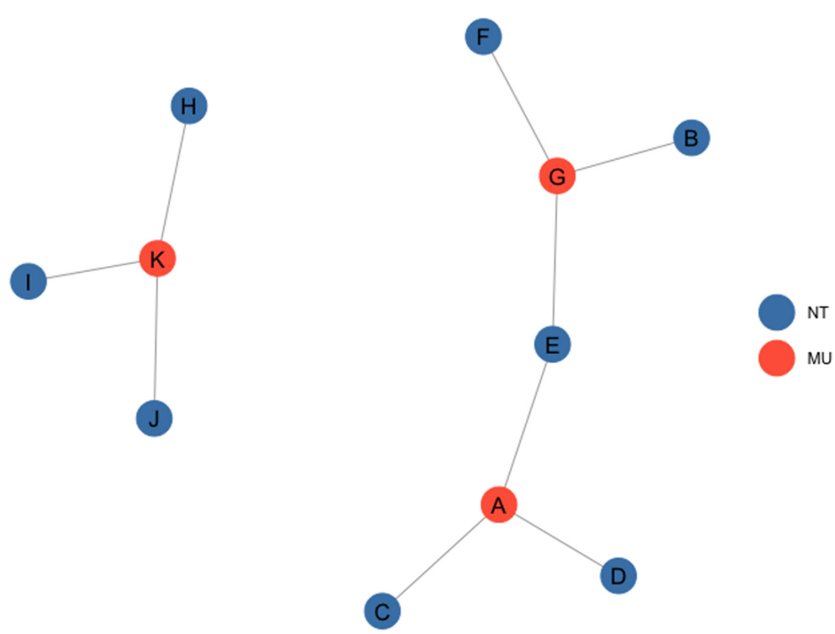

Figure S1: A hypothetical example of 1:3 matching of no-take (NT) and multi-use (MT) sites.

- **Step 5'**: Repeat the Step 1'-4' for B times. Compute the bootstrap variance estimator for the bias-corrected matching estimator $\hat{\tau}$ as the empirical variance of $\{\hat{\tau}_b^*\}_{b=1}^B$.

S5 Extension to unit-level treatment assignments

Although we focus on the case with cluster treatment assignments (i.e., sites within a cluster receive the same treatment), our framework extends readily to the case with unit-level treatment assignments (i.e., sites with a

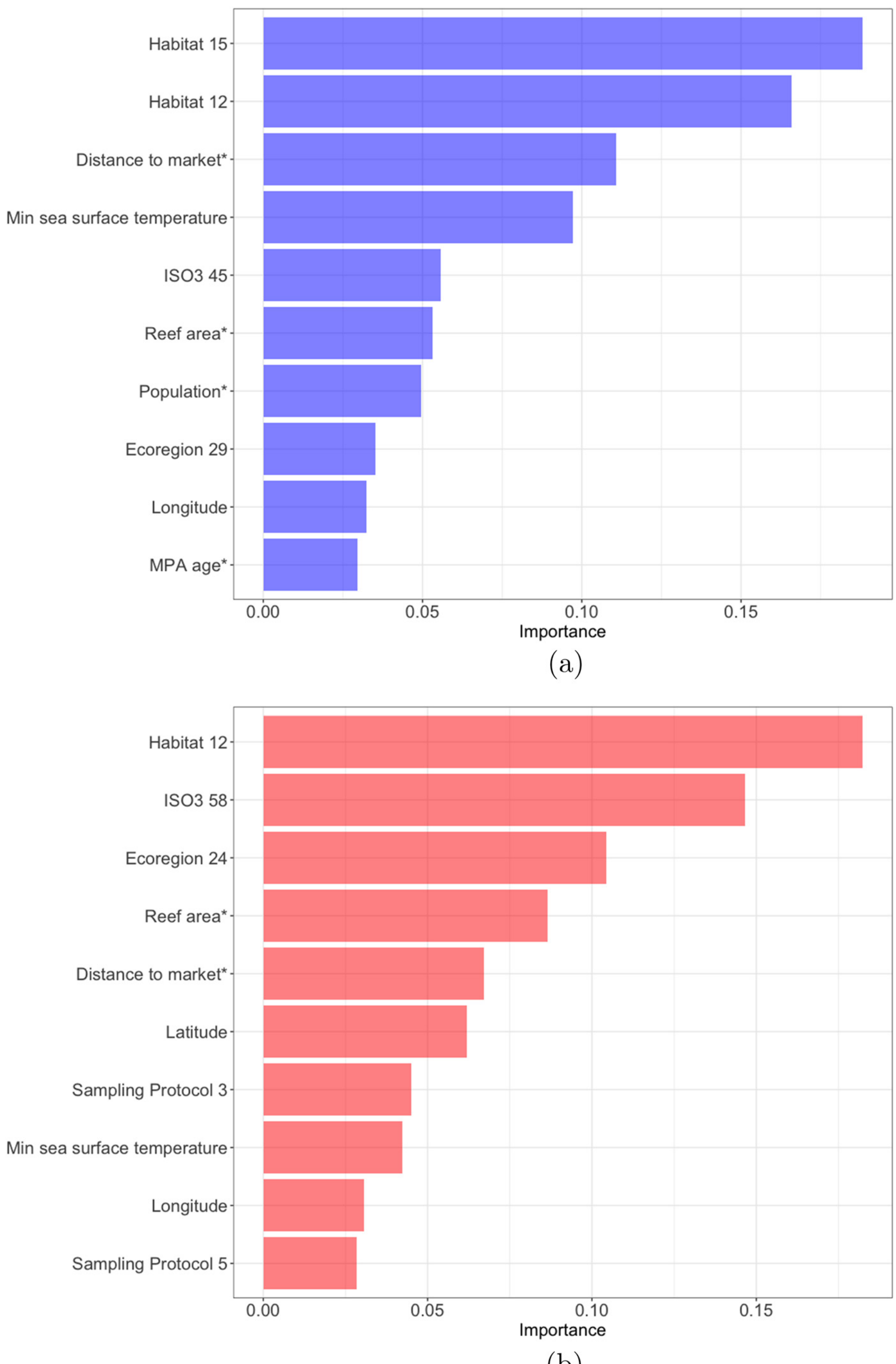

Figure S2: Top 10 important covariates under different marine protected area policies by regression forest (*indicates Box-Cox transformation for the covariate). (a) Multi-use and (b) no-take.

cluster can receive different treatments). In the motivating application, most MPAs have cluster treatment, but some MPAs have a few, although a small number of sites under a different policy from most sites. It is because some MPAs contain multiple zones with different regulations [5]. To accommodate this situation, it is sufficient to make simple modifications for applying the above method, theory, and inference. First, we change the cluster-level treatment A_r to the unit-level treatment A_{ir}. The ignorability assumptions for the ATE and ATT become the followings.

Assumption S4. (i) $\{Y_{ir}(0), Y_{ir}(1)\} \perp\!\!\!\perp A_{ir}|X_{ir}, Z_r$; (ii) $\eta < P(A_{ir} = 1|X_{ir} = x, Z_r = z) < 1 - \eta$ almost surely, for some $\eta > 0$.

Assumption S5. (i) $\{Y_{ir}(0)\} \perp\!\!\!\perp A_{ir}|X_{ir}, Z_r$; (ii) $P(A_{ir} = 1|X_{ir} = x, Z_r = z) < 1 - \eta$ almost surely, for some $\eta > 0$.

In the matching procedure, matching based on cluster-level confounders is insufficient to remove confounding biases. Thus we use both cluster-level and site-level confounders in matching, which is also recommended for the case with cluster treatment assignments. The role of site-level confounders removes confounding biases instead of improving the efficiency of the matching estimator.

S6 Supplementary figures

The letters A, G, and K represent sites under the multi-use policy, while the rest of the letters represent sites under the no-take policy. With 1:3 matching, one multi-use site is matched with three no-take sites. Meanwhile, the matched no-take sites can be paired with other multi-use sites. For example,the matched no-take site E is used twice to match site A and G (Figures S1 and S2)

References

[1] Abadie A, Imbens GW. Bias-corrected matching estimators for average treatment effects. J Business Econ Stat. 2011;29(1):1–11.
[2] Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. Econometrica. 2006;74(1):235–67.
[3] Serfling RJ. Contributions to central limit theory for dependent variables. Ann Math Stat. 1968;39(4):1158–75.
[4] Otsu T, Rai Y. Bootstrap inference of matching estimators for average treatment effects. J Am Stat Assoc. 2017;112(520):1720–32.
[5] Horta e Costa B, Claudet J, Franco G, Erzini K, Caro A, Gonçalves EJ. A regulation-based classification system for Marine Protected Areas (MPAs). Marine Policy. 2016;72:192–8.