S1 Appendix. Details of the statistical analysis

The original equation for an approximate design unbiased estimator of the mean temperature ($\hat{\mu}_{kd}$) for the range of seedlings or mature trees (d) of species k from Monleon and Lintz (2015) is given by the weighted domain sample mean:

$$\hat{\mu}_{kd} = \frac{1}{|\hat{D}_{kd}|} \sum_{i=1}^{n} I_{kd}(p_i) y(p_i) / \pi(p_i) \tag{1}$$

$$|\hat{D}_{kd}| = \sum_{i=1}^{n} I_{kd}(p_i) / \pi(p_i) \tag{2}$$

where $I_{kd}(p_i)$ is an indicator variable that takes the value of 1 if the i-th plot contains seedlings or mature trees of the k-th species or 0 if otherwise; $y(p_i)$ is the i-th plot value for temperature and $\pi(p_i)$ is the inclusion density function in the i-th plot, respectively; $|\hat{D}_{kd}|$ is an estimator of the size of the domain (range of seedlings or mature trees).

The equations in this appendix apply to both the species range analysis and the species density analysis. In this study $y(p_i)$ is the i-th plot value for temperature or precipitation (y) and d indexes one of two conditions being compared, P and Q. For the species range analysis P and Q represent plots occupied at the first visit and second visit respectively. For the species density analysis P and Q represent plots with fewer or more trees at the second visit. $I_{kd}(p_i)$ is an indicator variable that takes the value of 1 if the i-th plot meets condition P or Q and 0 if otherwise; $|\hat{D}_{kd}|$ is an estimator of the domain size (number of plots that meet condition P or Q).

~~In this analysis we conduct post stratification.~~ We divided the study region into 25 strata which have different sampling intensities. Because we are using stratified sampling, and inclusion probabilities are the same within strata for $\hat{\mu}_{kd}$ and $|\hat{D}_{kd}|$, we redefine equations 1 and 2 for specific strata as:

$$\hat{Y}_{kdh} = \sum_{i=1}^{n_h} I_{kdh}(p_i) y(p_i) \tag{3}$$

$$|\hat{D}_{kdh}| = \sum_{i=1}^{n_h} I_{kdh}(p_i) \tag{4}$$

where there are n_h plots in stratum h, with an overall total of n plots. The stratified means are defined (after equation 4.13 in Scott et al., 2005) as:

$$\hat{Y}_{kd} = \sum_{i=h}^{H} W_h \hat{Y}_{kdh} \tag{5}$$

$$|\hat{D}_{kd}| = \sum_{i=h}^{H} W_h |\hat{D}_{kdh}| \tag{6}$$

where H is the total number of strata and W_h is the area weight for the stratum. The area weight is not based on plot number, but rather on a predefined fraction of stratum area relative to the whole. The variance for $\hat{Y}_{kd}$ and $|\hat{D}_{kd}|$ can be given by the equation for an arbitrary estimated mean $\hat{G}$ (after equation 4.12 in Scott et al., 2005) as:

$$\rho(\hat{G}) = \frac{\sum_{i=1}^{n_h} y_{hi}^2 - n_h \hat{G}^2}{n_h (n_h - 1)} \tag{7}$$

We are interested in the difference in the ratio of trees of species k divided by their encounter probability $|\hat{D}_{kd}|$ for condition Q minus the same ratio for condition P. We obtain the ratio for condition P or Q using equations 5 and 6 which are based on equation 4.16 from Scott et al. (2005):

$$\hat{\mu}_{kd} = \frac{\hat{Y}_{kd}}{|\hat{D}_{kd}|} = \frac{\sum_{i=h}^{H} W_h \hat{Y}_{kdh}}{\sum_{i=h}^{H} W_h |\hat{D}_{kdh}|} \tag{8}$$

The variance for $\hat{\mu}_{kd}$ from equation 4.17 in Scott et al. (2005) is:

$$\hat{V}(\hat{\mu}_{kd}) \approx \frac{1}{|\hat{D}_{kd}|^2}\left[\hat{\nu}(\hat{Y}_{kd}) + \hat{\mu}_{kd}^2 \hat{\nu}(|\hat{D}_{kd}|) - 2\hat{\mu}_{kd}\hat{c}(\hat{Y}_{kd}, |\hat{D}_{kd}|)\right] \tag{9}$$

where $\hat{\nu}(\cdot)$ and $\hat{c}(\cdot,\cdot)$ are the estimated variance and covariance functions, respectively. For arbitrary estimators $\hat{x}$ and $\hat{z}$, they are, from equations 4.14 and 4.18 in Scott et al. (2005):

$$\hat{\nu}(\hat{x}) = \frac{1}{n}\left[\sum_{i=h}^{H} W_h n_h \rho(\overline{x_h}) + \sum_{i=h}^{H}(1 - W_h)\frac{n_h}{n}\rho(\overline{x_h})\right] \tag{10}$$

$$\hat{c}(\hat{x},\hat{z}) = \frac{1}{n}\left[\sum_{i=h}^{H} W_h n_h cov(\overline{x_h}, \overline{z_h}) + \sum_{i=h}^{H}(1 - W_h)\frac{n_h}{n}cov(\overline{x_h}, \overline{z_h})\right] \tag{11}$$

where, from equation 4.19 in Scott et al. (2005):

$$cov(\overline{x_h}, \overline{z_h}) = \frac{\sum_{i=1}^{n_h} x_{ih}z_{ih} - n_h\overline{x_h}\ \overline{z_h}}{n_h(n_h - 1)} \tag{12}$$

The latter portions of (10) and (11) are inconsequential, so (10) and (11) simplify to:

$$\hat{\nu}(\hat{x}) = \frac{1}{n}\left[\sum_{i=h}^{H} W_h n_h \rho(\hat{x}_h)\right] \tag{13}$$

and

$$\hat{c}(\hat{x},\hat{z}) = \frac{1}{n}\left[\sum_{i=h}^{H} W_h n_h cov(\overline{x_h}, \overline{z_h})\right] \tag{14}$$

For each species we estimated the difference in the mean annual temperature or precipitation for conditions P and Q as the difference between their respective domain ratio estimators:

$$\hat{\delta}_k = \hat{\mu}_{kQ} - \hat{\mu}_{kP} \tag{15}$$

The approximate variance of $\hat{\delta}_k$, using a Taylor series approximation (equation 6.9.1 in Wolter 1995), is:

$$\hat{V}\left(\hat{\delta}_k\right) \approx \hat{V}\left(\hat{\mu}_{kQ}\right) + \hat{V}\left(\hat{\mu}_{kP}\right) - \frac{2}{|\hat{D}_{kP}||\hat{D}_{kQ}|} \left[\begin{array}{c} \hat{c}\left(\hat{Y}_{kP}, \hat{Y}_{kQ}\right) + \hat{\mu}_{kP}\hat{\mu}_{kQ}\hat{c}\left(|\hat{D}_{kP}|, |\hat{D}_{kQ}|\right) \\ -\hat{\mu}_{kP}\hat{c}\left(\hat{Y}_{kQ}, |\hat{D}_{kP}|\right) - \hat{\mu}_{kQ}\hat{c}\left(\hat{Y}_{kP}, |\hat{D}_{kQ}|\right)] \end{array} \right] \quad (16)$$

We estimated a 95% confidence interval as:

$$\hat{\delta}_k \pm z_{0.975}\left[\hat{V}\left(\hat{\delta}_k\right)\right]^{1/2} \quad (17)$$

where $z_{0.975}$ is the 97.5 percentile of the normal distribution.

We obtained 95% confidence intervals using the percentile bootstrap method (Shao and Tu 1995). We sampled all plots with replacement to create a sample of the same size. We repeated the process 20,000 times to obtain bootstrap confidence intervals for the mean estimates.

Estimation of mean differences across all species

The estimated variances for individual species differed widely. Since estimators were calculated from the same sample and because tree distributions are non-random, the estimators are expected to be correlated. To estimate a mean difference across all species we dealt with the unequal and non-independent variances by using a generalized least-squares estimator:

$$\hat{\Delta} = \left(\mathbf{c^T}\mathbf{\Sigma^{-1}}\mathbf{c}\right)^{-1}\mathbf{c^T}\mathbf{\Sigma^{-1}}\hat{\boldsymbol{\delta}} \quad (18)$$

where $\hat{\boldsymbol{\delta}}$ is a vector of the individual species differences, $\hat{\delta}_k$, $\mathbf{c}$ is a vector of 1s of the same length as $\hat{\boldsymbol{\delta}}$, and $\mathbf{\Sigma}$ is the variance-covariance matrix of $\boldsymbol{\delta}$. The variance estimator is:

$$V(\hat{\Delta}) = \left(\mathbf{c^T \Sigma^{-1} c}\right)^{-1} \tag{19}$$

We estimated the variance-covariance matrix $\mathbf{\Sigma}$ by using output from the bootstrap procedure described above. We computed $\boldsymbol{\delta}$ of between-condition differences $(Q - P)$ and used the resulting matrices of all iteration outputs to obtain $\mathbf{\Sigma}$ for the mean estimate.

References

Fuller, W.A. 2009. Sampling Statistics. Hoboken: Wiley. 454 p.

Monleon, V.J., and H.E. Lintz. 2015. Evidence of tree species' range shifts in a complex landscape. PLoS One, 10(1), p.e0118069.

Särndal, C.-E., B. Swensson, and J. Wretman. 2012. Model assisted survey sampling. New York: Springer-Verlag Publishing.

Scott, C.T., W.A. Bechtold, G.A. Reams, W.D. Smith, J.A. Westfall, M.H. Hansen, and G.G. Moisen. 2005. Chapter 4: Sample-based estimators used by the Forest Inventory and Analysis National Information Management System. *In:* Bechtold, W. A. and P.L. Patterson, Editors, The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Shao, J., Tu D. 1995 The Jackknife and Bootstrap. New York: Springer-Verlag. 517 p.