

S1 Appendix. Details of the statistical analysis

The original equation for an approximate design unbiased estimator of the mean temperature for a species during the first or second visit ($\hat{\mu}_{kd}$) is given by the weighted domain sample mean (Monleon and Lintz 2015):

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

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

where k indexes species and d indexes whether the estimator is for trees at the first visit (F) or second (S); $I_{kd}(p_i)$ is an indicator variable that takes the value of 1 if the i -th plot contains first or second visit trees of the k -th species or 0 if otherwise; $y(p_i)$ and $\pi(p_i)$ are the value of temperature and the inclusion density function in the i -th plot, respectively; $|\hat{D}_{kd}|$ is an estimator of the size of the domain (range of the trees from the first or second visit); and the sum is over n , the total sample size.

In this analysis we will be conducting post stratification. We have 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:

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

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

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:

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

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

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 ($P1POINTCNT_h/p1pntcnt_{eu}$). 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:

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

We are interested in the difference in the ratio of trees of species k divided by their encounter probability $|\hat{D}_{kd}|$ encountered on the second visit minus the ratio for the first visit. We obtain the ratio for each visit using equations 5 and 6 using equation 4.16 from Scott et al. (2005):

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

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

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

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):

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

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

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

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

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

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

and

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

For each species we estimated the difference in the mean annual temperature for the range of trees at the first and second visit as the difference between their respective domain ratio estimators:

$$\hat{\delta}_k = \hat{\mu}_{kS} - \hat{\mu}_{kF} \quad (15)$$

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

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

We estimated a 95% confidence interval as:

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

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

We estimated the 5th and 95th percentiles of differences in temperature distribution between visits for each species by calculating the inverse of the empirical cumulative distribution function at 0.05 and 0.95 (Fuller 2009). 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 and computed the 5th and 95th percentiles of between-visit differences for each iteration. We repeated the process 20,000 times to obtain bootstrap confidence intervals for the mean and percentile 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:

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

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:

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

We estimated the variance-covariance matrix $\mathbf{\Sigma}$ by using output from the bootstrap procedure described above. We computed $\boldsymbol{\delta}$ along with the 5th and 95th percentiles of

between-visit differences and used the resulting matrices of all iteration outputs to obtain Σ for the mean and percentile estimates.

References

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

V.J. Monleon 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 programnational 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