

**A Bayesian approach to assessing impacts of invasive plants on native plant species richness in City of Austin parks  
SR-15-17, September 2015**

Angelita Rodriguez, Abel Porras, Ana González  
City of Austin  
Watershed Protection Department  
Environmental Resource Management Division

**Abstract**

*The Austin Invasive Management program conducted a data collection effort at various City of Austin parks during the summer of 2013 to evaluate the distribution and impacts of invasive plant species across the city. A Bayesian approach was used to characterize the native plant species richness in a park and to examine whether invasive species cover has an impact on native species richness in wooded habitats. This approach is useful when sample size is limited and useful prior information on species distribution is available. Furthermore, Bayesian methods provide an assessment on the probabilities of the hypotheses being true given the data rather than the probability of obtaining the data given various hypotheses being true. Although the analysis identified few parks where invasive plants have had a significant impact on native plant species richness, in most parks this was not the case. Potential dataset limitations were identified. Further research on specific invasive species is recommended.*

**Introduction**

The City of Austin adopted in 2012 the COA Invasive Species Management Plan (ISMP), which identifies 5-year invasive species management goals for all City departments that engage in vegetation management. The Austin Invasive Management (AIM) program forms the basis for implementation of the ISMP. In 2013, the AIM program focused on two of the ISMP goals: 1) the collection of baseline data on invasive species distributions on City-managed properties and 2) public education through the training and mobilization of volunteers in invasive species identification and data collection methods. In order to help meet these goals, the City applied for and was awarded a grant from the Bloomberg Philanthropies Cities of Service Impact Volunteering Fund for \$25,000. This grant allowed WPD to team with the Ladybird Johnson Wildflower Center to train 144 volunteers in identification and data collection methods for invasive plants. In addition, five interns were hired for the summer of 2013 to spearhead the sampling effort.

Watershed Protection Department staff collected baseline invasive species and native species data, as well as ecological information (i.e., community structure, habitat and soil types), for parks and greenbelts across the city during the summer of 2013. Initial data exploration examined whether native species richness was correlated to the percent cover of and/or the number of invasive species in each of the vegetation strata (groundcover, understory, and canopy). However, large variances in the data among sampled parks prevented detecting any correlation. Therefore, a stratified data analysis was utilized because it minimizes some of the variation, but at the expense of producing smaller data sets. Given its strength with small ecological data sets (McCarthy 2007, Johnson 1999, Anderson et al. 2000), a Bayesian approach was selected to characterize the impacts of invasive plants on the native plant species richness in each park across the city. This report will discuss the theory and the application of Bayesian inference, which will be followed by results from the data collection and conclusions and recommendations. Given the novelty of this approach to look at the impact of native and invasive species in Austin's parks, this report is meant as a "thought experiment" to inspire formulation and/or expression of potential mechanisms driving the distribution of invasive species in the City of Austin.

## **Methods**

A subset of the points identified as wooded habitat in the AIM dataset were used for this Bayesian approach. Site selection and data collection are documented elsewhere (Rodriguez et al. 2014). Wooded habitats were the primary focus for this analysis because they better represent the riparian vegetation communities typically found along streams. Wooded riparian habitats provide beneficial ecosystem services such as a reduction in nutrients and sediment input, pollutant removal, increase in groundwater infiltration, stabilization of stream banks, temperature buffering, carbon sequestration, plant and animal habitat (Fischer and Fischenich 2000, Jones et al. 2006, Richardson and Holmes 2007, Woolsey et al. 2007) as well as recreational opportunities.

The data used in this study includes, for each of the sampling plots, the percent cover of 13 invasive species identified in the Project Development Guide and the count of native species at each stratum of vegetation (groundcover, understory and canopy, (Rodriguez et al. 2014).

Parks with no wooded habitat plots (Slaughter Creek Park and Colony Park) were excluded from the analysis. In addition, some parks were excluded from the analysis for those strata for which the percent of invasive species was less than 3% or the number of plots with more than 3% was less than four (Walnut Creek Greenbelt 1, Upper Bull Creek Greenbelt 2, Tanglewood, Shoal Creek, Lower Bull Creek, Hielscher Tract, Circle C Ranch, Brushy Creek Greenbelt 2, and Brushy Creek Greenbelt 1).

Data stratification by park and vegetation stratum produced small sample sizes for some park/stratum combinations. To overcome this, a Bayesian approach was used to estimate the expected number of native species per plot in wooded habitats by park and by stratum (i.e., understory or groundcover) and assess the effects of invasive species cover on native species counts using three generalized linear models from the Poisson family of distributions.

The first model determined the expected count of native species independent of any covariate. Within each stratum (groundcover, understory, and canopy), the second model estimated the native species count dependent on the invasive species percent cover (e.g. the native species count in the understory dependent on the invasive species percent cover in the understory). The third model estimated the native species count dependent on the invasive species percent cover of a higher stratum (e.g. the groundcover native species count dependent on the invasive species cover on the understory).

## Theory

Bayesian statistical methods are derived from Bayes' Theorem which states that:

$$p(\theta|y) = p(y|\theta) \cdot p(\theta)/p(y) \quad (1)$$

In its most general form,  $\theta$  and  $y$  can be referred to as *events*, or a set of outcomes from an experiment. Thus, the theorem evaluates the probability of an event happening given the occurrence of the other event. When one considers  $\theta$  to be a set of unknown parameters<sup>1</sup> and  $y$  to be the data set, then Bayes' Theorem can evaluate the probability of a range of values from the set of unknown parameters,  $p(\theta|y)$ , due to the occurrence of the data,  $p(y|\theta)$  (Robert, 2007).

This is done by first specifying an assessment of the unknown parameters,  $p(\theta)$ , typically named as the prior distribution, or simply the *prior*. The assessment can be based on existing information or, if there is no information, then an uninformative or a large, prior distribution is used. This is often a source of controversy, but is actually a feature of Bayesian statistics because it can carry with it a distribution of values obtained by "experts". Any data collected afterward can be used by this approach to validate the priors if the experts are correct, or improve (or modify) the priors if the experts are incorrect (McCarthy, 2007). Thus, in either case, this information, which is not typically utilized in traditional statistical techniques, can inform (but not drastically change) the outcome of the analysis.

Next, the likelihood of the data set  $p(y|\theta)$  is determined, which is denoted as the *likelihood function*. Note that the likelihood of the data is calculated given an explicit assumption of the nature of the parameters. That is, a hypothesis or model of the parameters is assumed and made explicit. Thus, if one postulates that an unknown parameter (say, the mean of the population) comes from a standard normal distribution, then the likelihood of the data set can be computed by finding the value at  $y$  of the standard normal distribution. The joint probability of the data set can then be calculated from the intersection of all the data. This is performed over several values of the parameter (since the value of the parameter is unknown) to obtain the required probability density function of the data given the parameters.

Bayes' Theorem asserts that the probability of the unknown parameters given the data,  $p(\theta|y)$ , also referred to as the *posterior* distribution, is equal to the *likelihood function* of the data given the parameters,  $p(y|\theta)$ , multiplied by the *prior*,  $p(\theta)$ , divided by the normalizing constant  $p(y)$ . Once the posterior density is calculated, it can then be used as a prior density for future studies. It is important to state that the terms in Bayes' Theorem are probability distributions that require

---

<sup>1</sup> In this report, parameters will refer to statistical parameters, such as mean, standard deviation, or slope of a line from a linear regression.

the calculation of integrals (or summations for discrete probability distributions) over the range of possible values

A key feature of Bayesian statistics is in the explicit assumptions made for each distribution. The *priors* are declared clearly and can be expanded, if necessary. The *likelihood functions* can also be modified or tested, if there is some disagreement on the structure of the parameters. Or, as a corollary, several models of the parameters assumption can be tested and then compared to one another.

Furthermore, the prior and the likelihood function can be useful gauges of one another. If a large enough data sample is used, then the likelihood function would dominate and the prior distribution would become less relevant. However, if the data set is small, then the prior distribution has the potential to dominate the solution. If this prior distribution is informative, then the prior distribution tends to have a greater influence on the posterior distribution. To mitigate the influence of the prior, one aspires to update the priors as one continues gathering enough data to make a conclusive case on the values of the parameters.

Once the posterior distribution has been calculated, the 2.5 and 97.5 percentiles of the distribution can be obtained to generate 95% credible intervals of the parameter in question. These intervals indicate a 95% probability that the true value is within this range of values

## Analysis

In keeping with the key features of Bayesian inference, several priors and parameter models were used in the analysis. For all of the models, the counts of native species were used to calculate the likelihood of the data, based on the data coming from the unknown mean of a Poisson distribution. The following three models were used to describe the unknown mean:

1. The count of native species in the wooded habitats for each park and for each stratum is based on the mean count,  $\lambda$ , which is a random variable with no dependent variable. This will be designated the null model.
2. The count of native species in the wooded habitats for each park and within a stratum is based on the mean count, which is a random variable,  $\lambda$ , and linearly dependent on the percent cover of invasive species within the same stratum.
3. Within the wooded habitats, the count of native species in the groundcover of each park is based on the mean count, which is a random variable,  $\lambda$ , and linearly dependent on the percent cover of invasive species in the understory.

Thus, the models are all represented by the equation:

$$Counts \sim Pois(\lambda) \tag{2}$$

The term *Counts* refers to the actual number of species in each plot and it is a realization from a Poisson distribution. The parameter,  $\lambda$ , is the mean count and is used to characterize which Poisson distribution is used at each park. In the first model, there are no explanatory variables, and the mean is calculated based solely on the count data. In the second model, the mean count,  $\lambda$ , is dependent on one explanatory variable: the percent cover of invasive species in that stratum. In the third model, the mean count in the groundcover,  $\lambda$ , is dependent on one explanatory variable, but this variable is the percent cover of invasive species in the higher stratum

(understory). As an additional caveat to the second and third models, the mean count,  $\lambda$ , cannot be negative. Thus, the explanatory variables for the second and third models are raised to the exponential. That is represented by:

$$\lambda = e^{\alpha + \beta x} \quad (3)$$

Or an equivalent model:

$$\ln(\lambda) = \alpha + \beta x \quad (4)$$

In Equations (3) and (4),  $\alpha$  and  $\beta$  are the unknown parameters and  $x$  is a given value of the percent cover of the invasive species.

Equation (2) can be used to estimate the parameter,  $\lambda$ , for the first model. From this, one can compare parameter estimates of  $\lambda$  between parks. Equations (3) and (4) can be used to estimate  $\alpha$  and  $\beta$  for the second and third models. From the estimation of  $\beta$ , one can determine whether or not an effect from the percent cover of invasive species is impacting the count of native species in each park.

To conduct the statistical analysis, OpenBUGS software, which is a program for performing Bayesian analysis of statistical models using Markov chain Monte Carlo methods, was used to estimate  $\lambda$  and run the three models (Kéry 2010). Uninformative priors were used for the three models presented here. Posterior distributions of each model parameter were generated by executing and saving the 10,000 Monte Carlo Markov chain samples in each model.

## Results

The statistical analysis of the data collected is presented below. First, though, to fully understand the meaning of the parameters estimated, the analysis and results from one of the parks, Walnut Creek Copperfield, using the first model will be presented.

### Null Model

The following two priors were used for the first model: a gamma distribution with shape and scale parameters both equal to 0.01 (Figure 1) and a gamma distribution with shape and scale parameters equal to 7 and 0.5 (Figure 2), respectively. The gamma distribution is typically recommended (McCarthy 2007) as conjugate prior distributions for Poisson distributions. The two different priors represent differing expert opinions on low mean count (Figure 1) and higher mean count (Figure 2) of native plant species per plot.

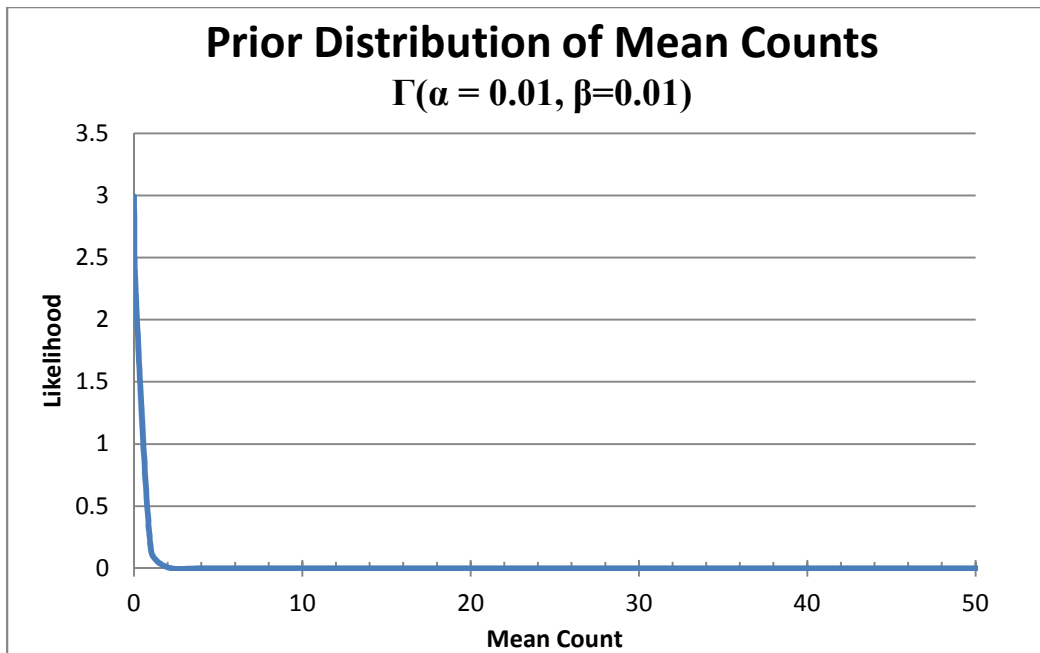


Figure 1: Gamma distribution with  $\alpha = 0.01, \beta=0.01$

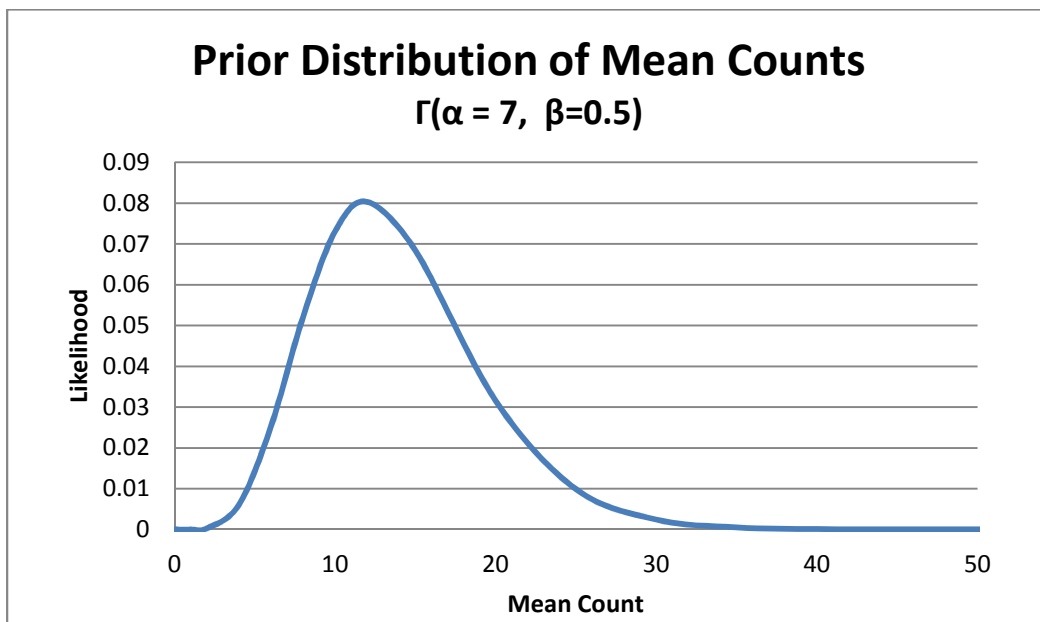


Figure 2: Gamma distribution with  $\alpha = 7, \beta=0.5$

For each of these two *priors*, OpenBUGS randomly<sup>2</sup> picks a value of  $\lambda$  based on the prior. For the first *prior*, OpenBUGS would likely pick a value less than 2. For the second *prior*, OpenBUGS would likely pick a much larger value. Then, each of these values is evaluated as a potential value for the mean count,  $\lambda$ . For each value of  $\lambda$ , a likelihood that each of the counts comes from the Poisson distribution with mean  $\lambda$  is computed and saved in computer memory. Based on the current value of  $\lambda$ , another value of  $\lambda$  is chosen<sup>3</sup>, its likelihood computed and saved.

<sup>2</sup> using Monte Carlo methods

<sup>3</sup> using Markov Chains processes

This process continued on for 10,000 iterations. This results in 10,000 values of  $\lambda$ , each with its own likelihood. Together, these values comprise the posterior distribution of  $\lambda$ . Figure 3 below shows the resulting posterior distributions from the two priors given above at Walnut Park Copperfield. The difference between the two posterior distributions is about 1 mean count per plot. This is not much considering the vast difference between the two priors. Furthermore, the resulting posterior distributions have a much smaller variance than the priors.

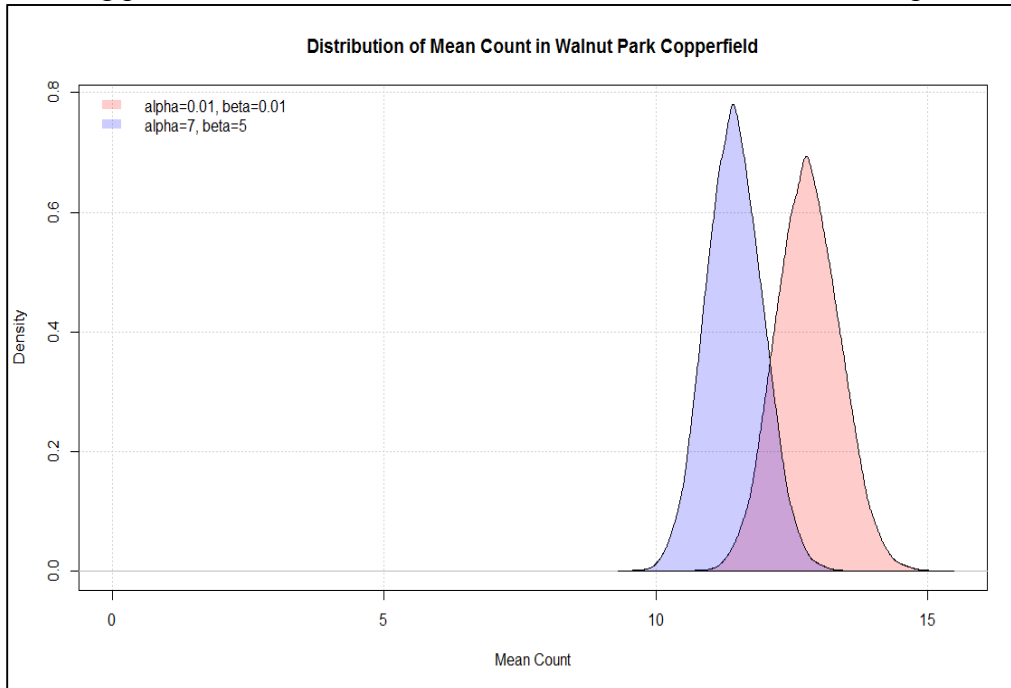


Figure 3: Posterior distributions of  $\lambda$  (mean count) for Walnut Park Copperfield from priors generated using the two gamma distributions ( $\alpha=0.01$ ,  $\beta = 0.01$  and  $\alpha = 7$ ,  $\beta = 0.5$ )

Varying the priors in this manner was performed for each park, and resulted in similar posteriors. This indicates that, for each park, the data dominates the solution, which indicates that the solution is mostly informed from the data rather than the prior. The resulting posterior distributions for all parks are displayed in Figure 4 using the higher mean count priors.

For most of the parks, the estimated range of the mean count of native species per plot found in the groundcover was between 10 and 14; Southpark Meadows had the greatest estimated mean count of native species per plot with a credible interval between 14 and 18. In contrast, Little Walnut Park and East Boggy Creek Greenbelt had the lowest estimated mean counts of native species in the groundcover with credible intervals between 6 and 7 counts and 9 and 11 counts, respectively. These estimates provide a baseline mean count of the native species per plot in the groundcover in each park and the expected range in the mean count.

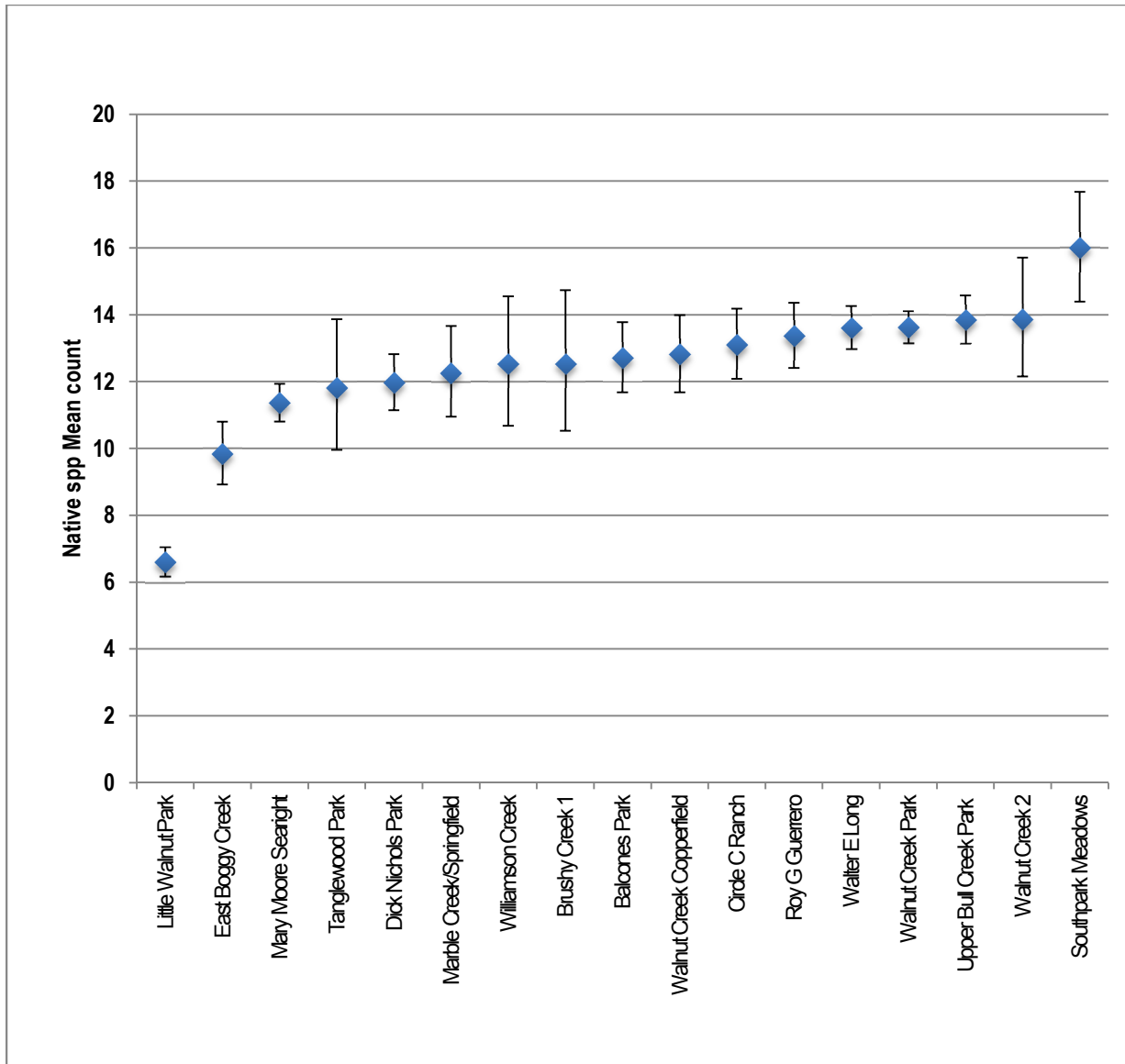


Figure 4. Estimated mean count and credible intervals of native species per plot for wooded habitats in the groundcover by park. Points represent the median value of the mean count are shown with error bars representing the 2.5 and 97.5 percentiles.

The estimated mean count per plot of native species found in the understory of parks was generally lower than in the groundcover at between 6 to 10 species (Figure 5). Walnut Creek Greenbelt 2 had the greatest number of native species found in the understory at between 9 and 12. Whereas, Little Walnut Park had the lowest mean number of native species (between 3 and 4). Overall, the groundcover has a higher diversity of natives than the understory. This result is not surprising, groundcover vegetation includes herbaceous species as well as the early recruitment of both understory and canopy species in the early stages of their life cycle.

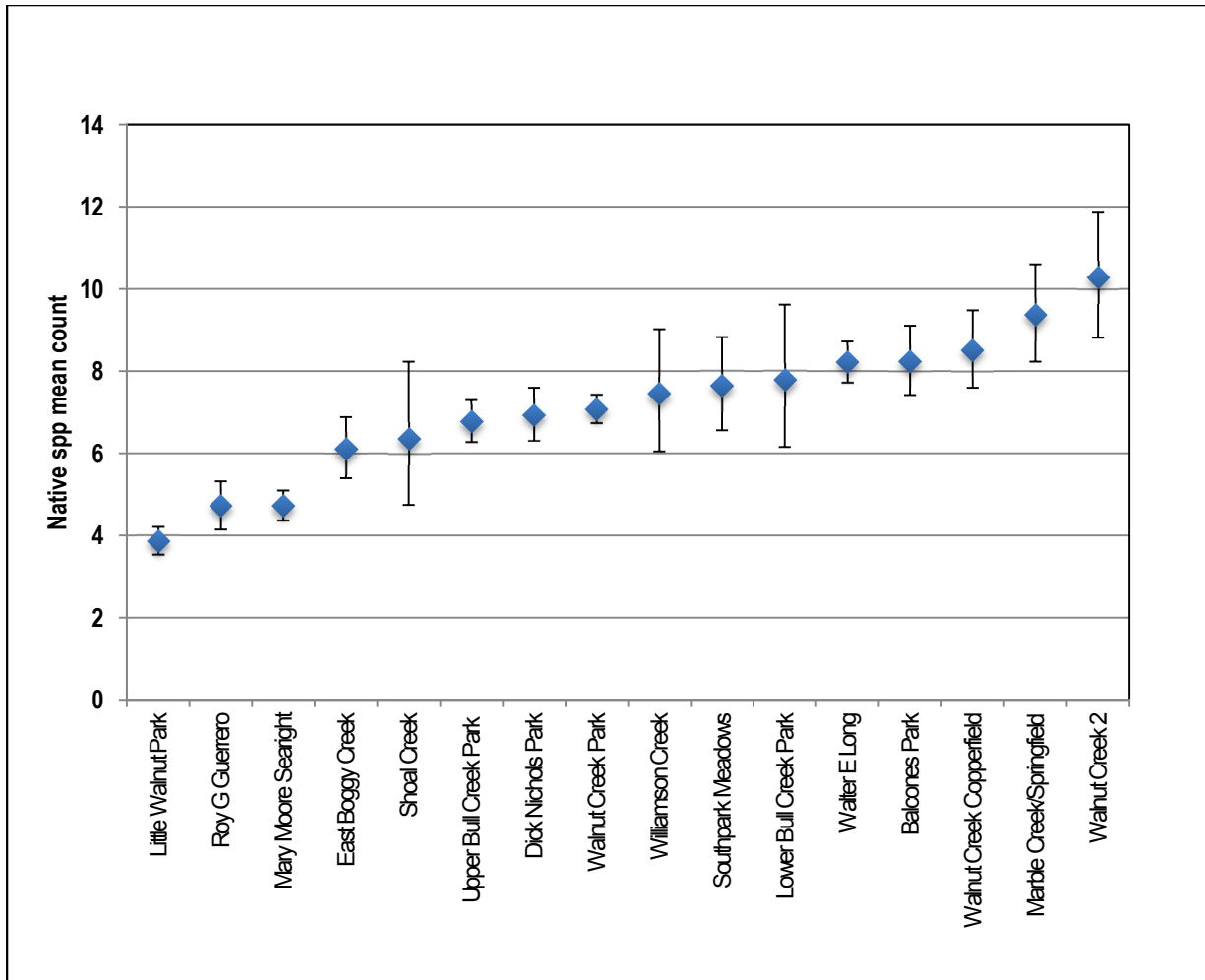


Figure 5. Estimated mean count and credible intervals of native species per plot for wooded habitats in the understory by park. Points represent the median value of the mean count are shown with error bars representing the 2.5 and 97.5 percentiles.

### Linear Models

The next step in the analysis of the data was to evaluate the variance of the posterior distributions for each park and examine whether some of this variation in mean native species count could be explained by the percent cover of the invasive species within those plots. We examined the potential effect of invasive species percent cover at two levels: within stratum and between strata:

1. The native species mean count in the groundcover stratum is linearly dependent on the percent cover of invasive species in the groundcover.
2. The native species mean count in the understory stratum is linearly dependent on the percent cover of invasive species in the understory.
3. The native species mean count in the groundcover stratum is linearly dependent on the percent cover of invasive species in the understory.

The canopy stratum was not examined in this analysis because the sample size was too small in most parks (17 of the 24 parks had less than four plots with at least 3% cover of invasive species). In addition, the effect of invasive species percent cover in the groundcover on the mean

native species count was not assessed; the underlying assumption is that light competition would be the driving exclusion mechanism in this system and therefore individuals of invasive species in the groundcover stratum did not have a strong effect on individuals of native species in the understory stratum.

The parameters for the linear models, as described in the Analysis section, are comprised of  $\alpha$  and  $\beta$ , which relate to the intercept and slope of the linear model, respectively. The intercept of the linear model can be interpreted as a mean count of native species and the slope as the change in mean counts of natives per the change in percent cover of invasive species. An estimate of the mean count (i.e. the intercept of the model) of native species for each park ranged from  $\ln(3)$  to  $\ln(16)$ <sup>4</sup>. Similarly, an estimate of the slope of the model can be inferred from the estimate of the mean counts divided by at least 10, since the percent invasive cover ranges from 10 to 100. Thus, our estimate of the slope of the model would at most be  $\pm 0.1$ . From this, an initial estimate of the priors for these parameters can be inferred to be normal distributions with a mean of zero and a variance of 1000 (roughly a standard deviation of 31). These priors were input into OpenBUGS and should more than adequately encompass the model estimates for  $\alpha$  and  $\beta$ .

#### Within Stratum: Groundcover

For most parks, the invasive species cover in the groundcover did not have a measurable impact on the native species counts within the same stratum. Impacts significantly different from zero (Figure 6, blue points with credible intervals not overlapping zero) were only present in four parks and were either negative (Mary Moore Searight and Little Walnut Park) or positive (Circle C Ranch and Balcones Park) depending on the park. The species with the highest cover values, and therefore the driver for these relationships, is glossy privet (*Ligustrum lucidum*), except at Circle C where Chinaberry (*Melia azedarach*) is the main invasive species. For all parks except one (East Boggy Creek Greenbelt) the range of invasive species cover was very limited. It is possible, therefore, that a failure to detect a significant relationship between invasive species cover and native species richness in most parks was a limitation of the dataset. Furthermore, it is possible that effects of invasive species cannot be detected unless a certain threshold of invasive species cover has been reached, as has been documented in other systems (e.g. Grice 2004, Gooden et al. 2009). However, in East Boggy Creek Greenbelt, the invasive species cover in the groundcover ranged from 0 to 80% and the estimated impact was not credibly different from zero.

---

<sup>4</sup> Recall from equation (4) that the counts are transformed by the natural logarithm.

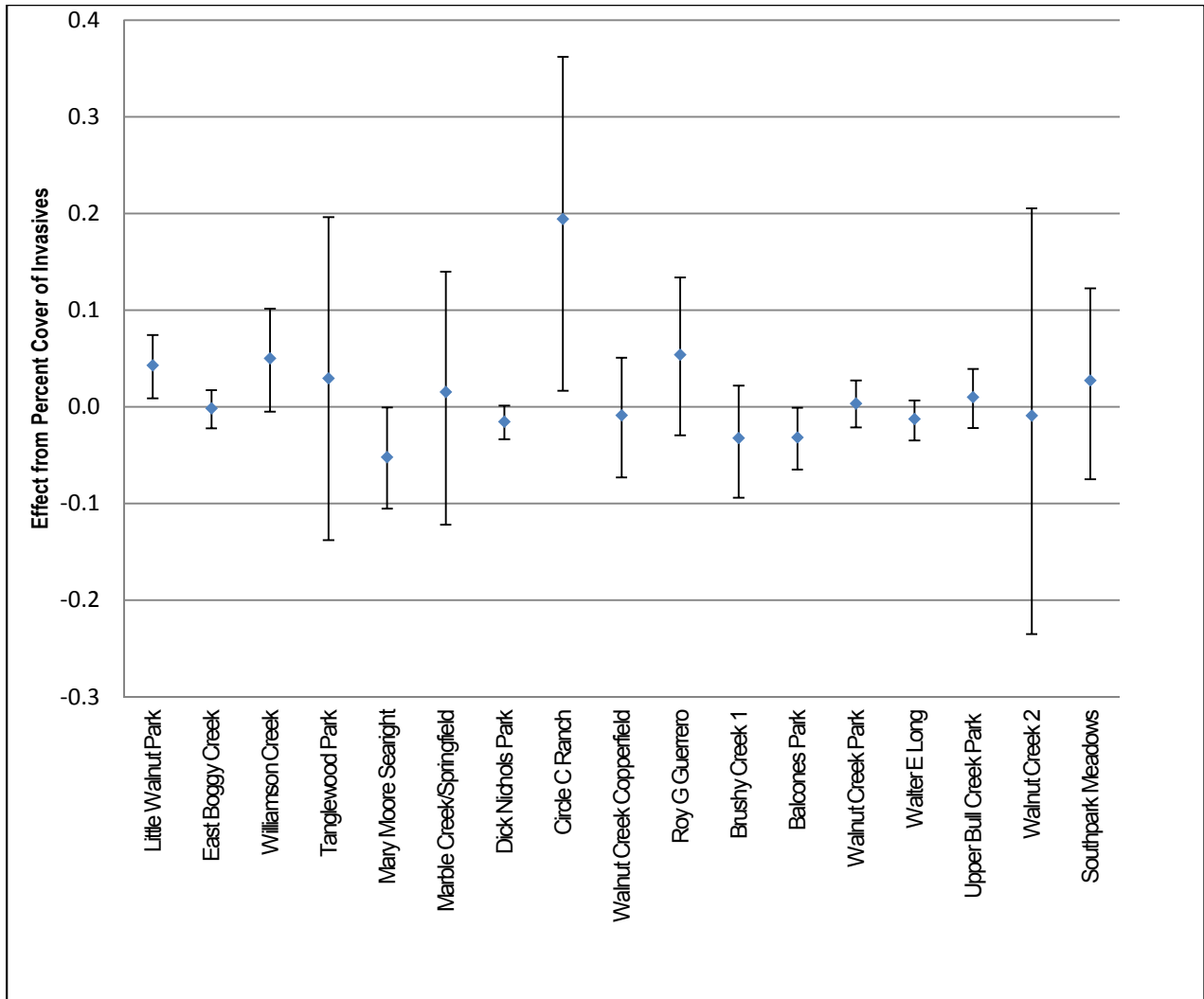


Figure 6. Estimated impact of invasive species percent cover in the groundcover stratum on the native species richness in the same stratum.

#### Within Stratum: Understory

In most parks (10 out of 16) there was no detectable impact of percent invasive species cover in the understory on native species counts within the same stratum. In East Boggy Creek Greenbelt, it had a negative effect, whereas in five parks (Mary Moore Searight, Little Walnut Park, Walnut Creek Park, Walnut Creek at Copperfield, and Upper Bull Creek Park) invasive species percent cover had a positive relationship to native species richness (Figure 7). The positive effect found in some of the parks is consistent with other research findings in which there are positive correlations between the abundance of exotic invasive species and native species diversity potentially explained by favorable conditions for species survival, regardless of origin (Levine and D’Antonio 1999). It is important to highlight that the range of the percent cover from the invasive species was limited to a small range that rarely exceeded 30%. Therefore, estimates of impacts from the linear models would not be valid above 30% percent cover of invasive species. If the effects of invasive species cover on native species are only detected beyond a certain threshold, a dataset including data points past that threshold would be needed to detect the impacts. East Boggy is the only park where the invasive species cover reaches up to 95% in

some plots and thus likely to be within a threshold at which the intensity of competitive interactions reach a level at which they are statistically detectable.

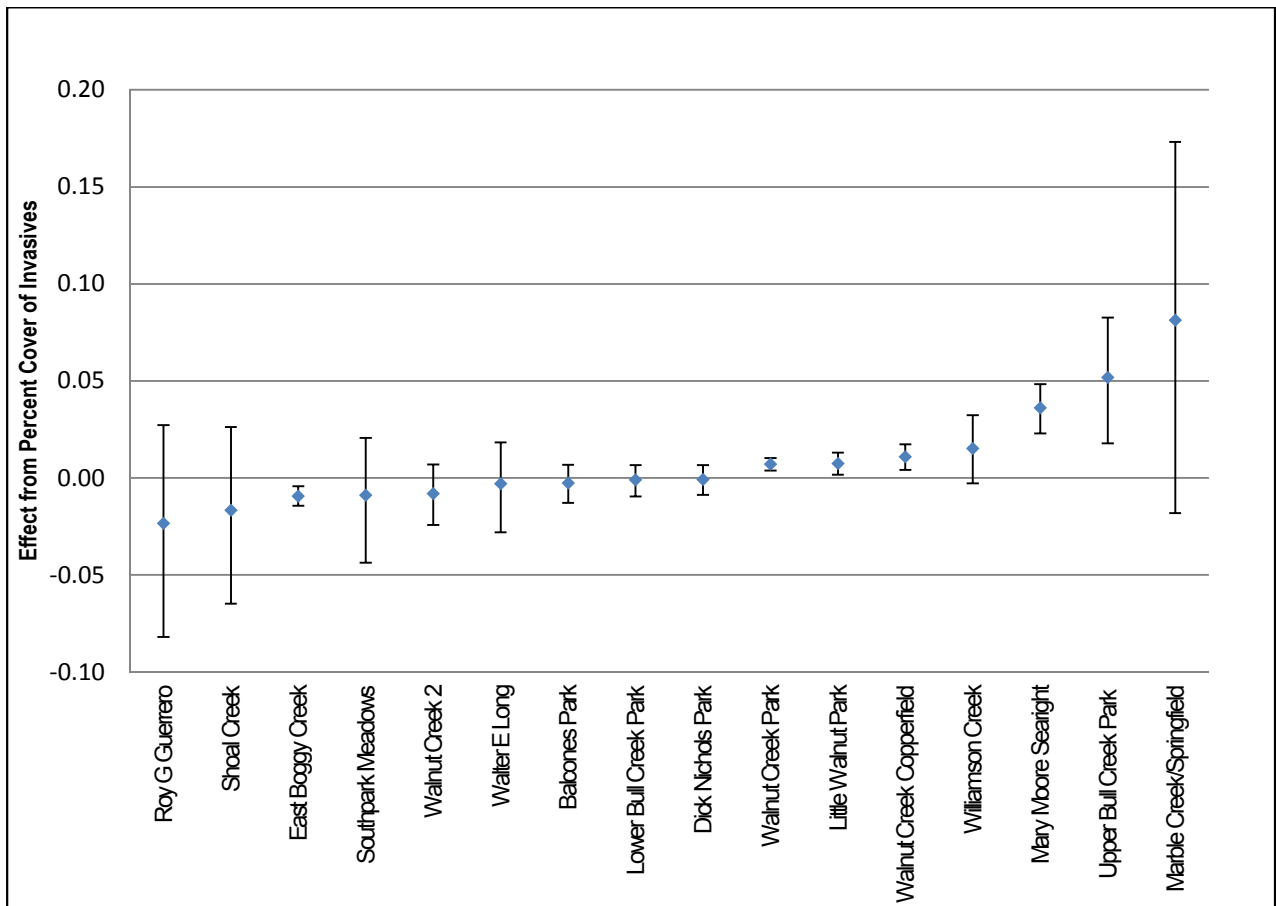


Figure 7. Estimated impact of percent cover of invasive species in understory on the native species found in the understory by park.

Between Strata Linear Model: Understory on groundcover

The third model examined whether there is an effect of the percent cover of understory invasive species on the native species richness in the groundcover. The effect of invasive species cover in the groundcover on the native species count in the understory was not examined as it was assumed plants mostly affect other plants within the same stratum or the stratum below through light competition (Royo and Carson 2006). Small sample sizes for invasive species cover in the canopy precluded examining its effect on native species numbers on the strata below (understory and groundcover).

Only three parks showed an effect from understory invasive species cover on the native species diversity in the groundcover. In Dick Nichols Park the impact was negative while it was positive in Mary Moore Searight and Lower Bull Creek Greenbelt (Figure 8). Glossy privet (*Ligustrum lucidum*) was the most abundant invasive species in the understory at Mary Moore Searight and

Lower Bull Creek Greenbelt. In Dick Nichols Park the invasive species dominating the understory was heavenly bamboo (*Nandina domestica*).

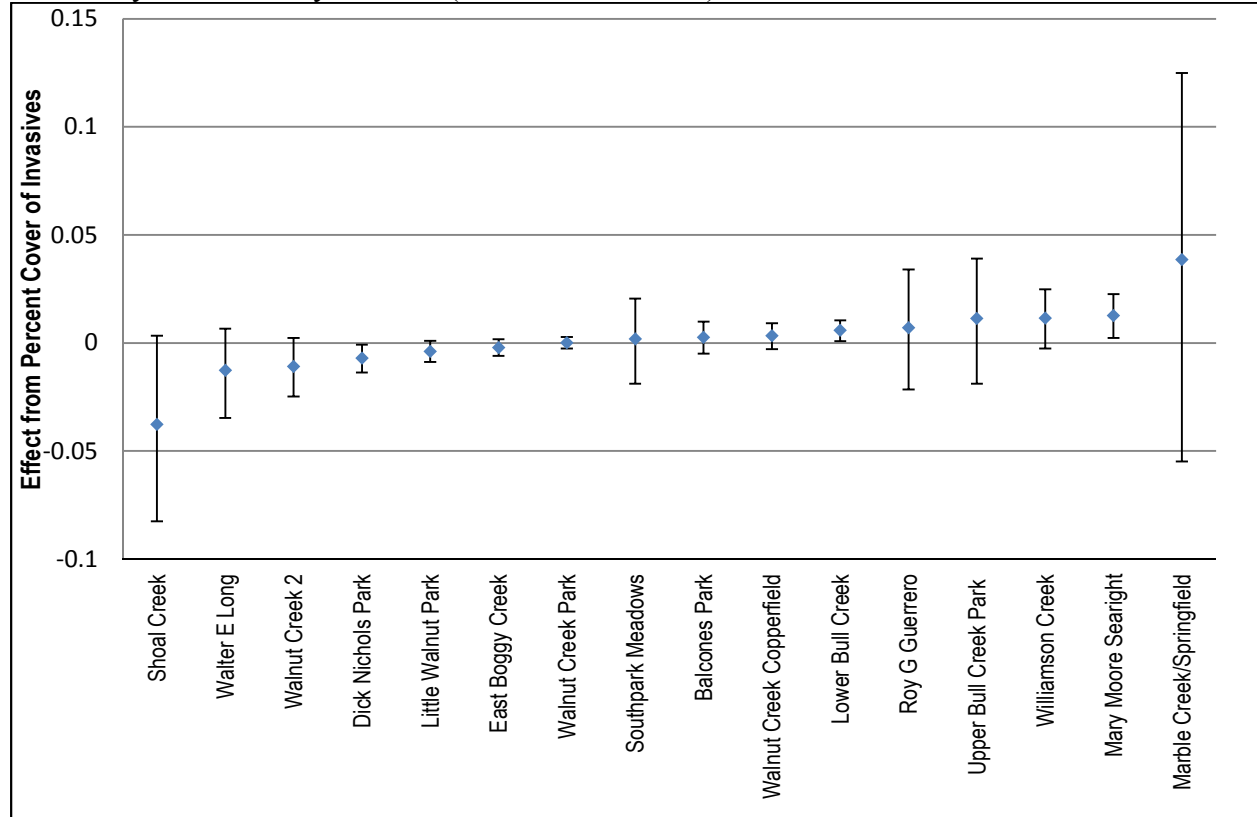


Figure 8. Estimated impact of percent cover of invasive species in understory on the native species found in the groundcover by park.

## Conclusions and Recommendations

This report presented analyses and results from an invasive species survey conducted in the summer of 2013. A simple null model provided basic counts of native species in the groundcover and understory for each park. Generally speaking, the groundcover consisted of a greater number of native species per plot than the understory. As discussed before, the understory consists of woody species whereas the groundcover contains both herbaceous species and seedlings of woody species. Therefore, the groundcover may be intrinsically more diverse as it contains individuals that complete their life cycle within the groundcover as well as those who will grow out of the groundcover and into the understory as saplings.

Linear models examined whether native species in each park are being impacted by percent cover of invasive species. In some parks a detectable impact of invasive species cover on native species was found. However, this effect was sometimes negative and sometimes positive. In most parks an effect was not detected. The limited range of invasive species cover in most parks precludes making inferences about these effects across the whole range of invasive species cover. Further data collection encompassing sites with a broader range of invasive plant cover for a focused set of species - *Ligustrum lucidum*, *L. sinense*, *L. quihoui*, and *Nandina domestica* - is recommended to assess the effect of these species with more confidence. One of the strengths of Bayesian inference is that the 95 percent credible intervals generated as a result of this study can function as priors to future studies, which would strengthen future analysis.

In addition, differences in other factors among parks may influence the distribution and interactions of invasive and native species. If this is the case, then analyses incorporating these factors into the Bayesian analysis would help explore additional factors such as presence of abiotic gradients (e.g. moisture) or distance to disturbance elements (roads, development) that may be informative. This report recommends exploring those potential spatial factors and/or additional environmental covariates in future analyses.

## Reference

- Anderson, D. R., K. P. Burnham, W. L. Thompson. 2000. Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management* 64:912-923.
- Fischer, R., and J. Fischenich. 2000. Design recommendations for riparian corridors and vegetated buffer strips. Ecosystem Management and Restoration Research Program. US Army Corps of Engineers:1-17.
- Gooden, B., K. French, P. Turner, and P. Downey. 2009. Impact threshold for an alien plant invader, *Lantana camara* L., on native plant communities. *Biological Conservation*:1-49.
- Grice, A. 2004. Weeds and the monitoring of biodiversity in Australian rangelands. *Austral Ecology* 29:51-58.
- Johnson, D. H. 1999. The Insignificance of Statistical Significance Testing. *The Journal of Wildlife Management* 63:763-772.
- Jones, A., P. Stacey, J. Catlin, and L. Stevens. 2006. A Rapid Stream-Riparian Assessment Protocol and its Utility in the Grand Staircase Region, Utah. [wildutahproject.org](http://wildutahproject.org).
- Kéry, M. 2010. Introduction to WinBUGS for Ecologists Chapter 4. Introduction to WinBUGS for Ecologists. First. American Press, Amsterdam.
- Levine, J., and C. D'Antonio. 1999. Elton revisited: a review of evidence linking diversity and invasibility. *Oikos* 87:15-26.
- McCarthy, M. 2007. *Bayesian Methods for Ecology*. Cambridge University Press, Cambridge, United Kingdom.
- Richardson, D., and P. Holmes. 2007. Riparian vegetation: degradation, alien plant invasions, and restoration prospects. *Diversity and Distributions* 13:126-139.
- Robert, C.P. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Verlag, New York.

- Rodriguez, A., D. Krenzelok, R. Clayton, A. Duncan, M. Scoggins, A. Gonzalez, and J. Clement. 2014. Austin Invasive Management (AIM) Project Development Guide. SR-14-05. City of Austin, Watershed Protection Department, Austin (TX).
- Royo, A., and W. Carson. 2006. On the formation of dense understory layers in forests worldwide: consequences and implications for forest dynamics, biodiversity, and succession. *Canadian Journal of Forest Research* 36:1345–1362.
- Woolsey, S., F. Capelli, and T. Gonser. 2007. A strategy to assess river restoration success. *Freshwater Biology* 52:752–769.