Lundbom Spotted Knapweed Data Analysis

Contract Number V17HRA505

Prepared for: Susan Turner, Biocontrol Specialist
Ministry of Forests Lands and Natural Resource Operations
Range Branch - Kamloops
441 Columbia Street
Kamloops BC V2C 2T3

Prepared by Julia R. Chandler
March 31st, 2017

Index

Introduction

Significant investment has been allocated towards the control of invasive weed species in the northwestern regions of the United States and Canada (Roger et al., 1998). Spotted knapweed (*Centaurea maculosa*), one such introduced pioneering plant species has been targeted by land managers to reduce the economic (Hirsch, 1996) and environmental (Lacey et al., 1989) impacts associated with its widespread dominance (DiTomasso, 2000). However, the use and effectiveness of insect biological control is under question (Mueller-Scharer and Schroeder, 1993; Callaway et al., 1999; Story et al., 2008).

The purpose of this report is to investigate how the introduction of five biological control insects (*Larinus minutus, Metzneria paucipunctella, Urophora affinis, Urophora quadrifasciata* and *Chaetorellia acrolophi*) may have influenced the seed and vegetative properties of spotted knapweed after more than 4 decades of an insect biological control program in the Lundbom Commonage near Merritt, British Columbia, Canada. Seed heads, soils and vegetation were sampled in 2016 at 3 sites dominated by spotted knapweed and the following is the preliminary investigation using the data from the 2016 sampling.

Statistical analysis reported herein include: (1) occupancy of seed heads by each biological control insect in the study by comparing the relationships between insect density per seed head; (2) tests for differences in counts of viable and non-viable seeds per seed head with insect density; and (3) tests for the geographic variation in seed bank, insect density, seed head seed viability and vegetation attributes.

Methods

Data Manipulation

All data manipulation and statistical analyses were performed using the programming language R (Team RDC 2016). For 'Attachment 4 Lundbom Knapweed Seed Head Data 2016.xlsx', the sites were appended to create a single spreadsheet with 151 observations x 52 variables and saved as '2016_seed head append.csv'. With script '*R code_DATA_restructure seed head data*' the data of '2016_seed head append.csv' were reshaped to long format thereby affording each observation (row) as a seed head sample, producing '2016_seedhead.csv' with 755 observations x 14 variables. 'Attachment 3 Lundbom knapweed veg data 2016.xlsx' and 'Attachment 10 Seed Soil Data 2016.xlsx' were joined by sample location (m) along each respective transect ('2016_seedsoil&veg.csv' with 62 observations x 14 variables).

With script '*R code_DATA*', '2016_seedhead.csv' and '2016_seedsoil&veg.csv' were imported and cleaned by removing incomplete cases due to sampling error (such as the offset of transects or loss of samples). The final resulting data frames were generated: 'seedhead' with 740 observations and 15 variables (an added variable for insect total) and 'vegsoil' 57 observations and 23 variables (an added variable for % viable seed along with soil sample and vegetation quadrat standardized (m^2)). There are an additional two matrices (each with 27 observations and 25 variables) that include the samples collected for seed head, soil sample and vegetation record for the same location (m) on each transect (at each site) that were aggregated either by the sum or the mean of the observations. These matrices are included with this report as .csv along with the R script described above.

Statistical analysis

Correlation matrices were constructed (Appendix A) to provide and initial overview of the structure of each variable and provide an indication of how all variables relate to each other. There are few direct correlations ($p < 0.05$) between the original (i.e., directly measured) variables with the exceptions of (1) non-viable seeds in the seed head, the seed bank, and germinated seed against ruptured seeds; (2) density of stems and rosettes against density of bolted plants; (3) *Urophora affinis* against *Larinus minutus* and density of rosettes ($n = 27$); and (4) *Urophora affinis* against *Larinus minutus* and *Urophora quadrifasciata*, as well as between *Larinus minutus*, viable seed in the seed head and non-viable seed head seed ($n = 740$).

The data are generally left-skewed and zero-inflated. This presents challenges when pursuing hypothesis testing and non-parametric measures (i.e., rank sum tests) were included in the analysis. To illustrate the co-occurrence of insect species in a single seed head, a venn diagram was constructed with the `draw.quad.venn()` function from the 'VennDiagram' package. Omnibus tests for differences in means by categorical independent variables (with their respective post-hoc analyses) were carried out with either of two methods:

1. `kruskal.test`() function and then post-hoc multiple comparisons with Conover's pairwise comparisons test by the `posthoc.kruskal.conover.test` () function from the 'PMCMR' package, while also including the Bonferroni correction for multiple comparisons. Additional non-parametric (Kruskal-Wallis rank sum) tests were conducted to determine differences in seed bank and vegetation parameters using a subset dataset that includes all variables ($n = 27$) but no significant differences in means were found between levels of insect density (not reported); or
2. `aov`() function followed by the `TukeyHSD`() function.

Results

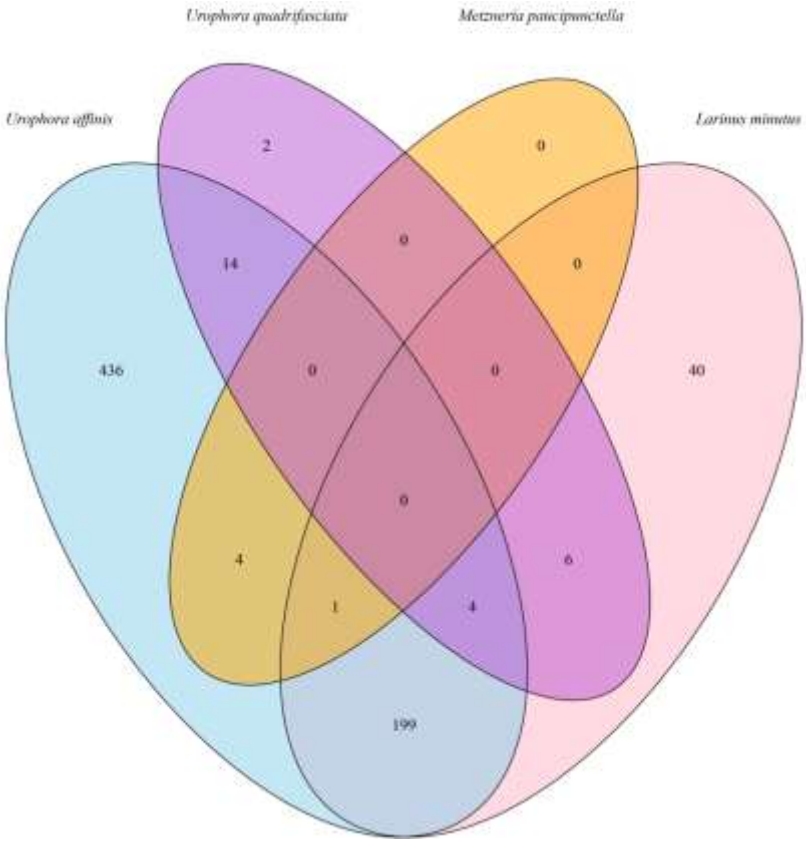


Figure 1. Venn diagram representing seed head occupancy (presence) of *Larinus minutus*, *Metzneria paucipunctella*, *Urophora affinis* and *Urophora quadrifasciata*. *Chaetorellia acrolophi* was not recorded in the 740 seed heads sampled. The maximum number of individuals/seed head varied by species: *Larinus minutus* (max. 2); *Metzneria paucipunctella* (max. 1); *Urophora affinis* (max. 10); and *Urophora quadrifasciata* (max. 4).

Variation in insect density (n = 740).

	*Urophora affinis**		*Larinus minutus**	
*Insect no.	*Larinus minutus*	*Urophora quadrifasciata*	*Urophora affinis*	*Metzneria paucipunctella*
0	0.6 ± 0.5 ^{a}	0.2 ± 0.6 ^{a}	3.3 ± 1.9 ^{a}	0.0 ± 0.1 ^{a}
1	0.5 ± 0.5 ^{a}	0.0 ± 0.2 ^{ab}	2.2 ± 1.7 ^{b}	0.0 ± 0.0 ^{a}
>1	0.3 ± 0.5 ^{b}	0.0 ± 0.2 ^{b}	2.6 ± 2.6 ^{ab}	0.1 ± 0.4 ^{b}
Kruskal-Wallis chi-squared	37.15	12.36	56.59	18.43
P	< 0.001	0.0021	< 0.001	< 0.001

Urophora affinis: 0 insects n = 82; 1 insect n = 103; >1 insect n = 555.
Larinus minutus: 0 insects n = 490; 1 insect n = 242; >1 insect n = 8.
Kruskal-Wallis rank sum tests; p < 0.05.
Pairwise comparisons made with Conover's test and the Bonferroni p-value adjustment.

Table 1. Kruskal-Wallis rank sum tests indicate differences in the mean density of (1) *Larinus minutus* and *Urophora quadrifasciata* with the mean density of *Urophora affinis*; and (2) the mean density of *Urophora affinis* and *Metzneria paucipunctella* with the mean density of *Larinus minutus.* Greater incidences of *Larinus minutus* and *Urophora quadrifasciata* were found in seed heads where *Urophora affinis* was absent compared to 2 or more individuals. Fewer incidences of *Urophora affinis* occurred in seed heads without *Larinus minutus* and *Metzneria paucipunctella* presence increased in seed heads with 2 or more *Larinus minutus* individuals.

Table 2. ANOVA with Tukey multiple comparisons of means to detect differences in counts of viable and non-viable seeds per seed head with insect density. Results suggest that *Larinus minutus* was the only insect associated with significant differences in the means of seed head seed counts (black and other). Results indicate reduced viable (black) and non-viable (other) seeds in the presence of *Larinus minutus*.

Variation in reproductive potential (seeds/seed head) of spotted knapweed (mean ± SEM) with insect density (n = 740).

Insect	*Larinus minutus*		*Urophora affinis*		*Urophora quadrifasciata*		*Metzneria paucipunctella*	
no.	black	other	black	other	black	other	black	other
0	2.5 ± 0.2 ^{a}	4.2 ± 0.2 ^{a}	1.9 ± 0.3	3.1 ± 0.5	2.2 ± 0.1	3.6 ± 0.2	2.2 ± 0.1	3.6 ± 0.2
1	1.7 ± 0.2 ^{b}	2.4 ± 0.2 ^{b}	2.3 ± 0.4	3.5 ± 0.5	3.4 ± 1.0	2.8 ± 0.8	3.2 ± 1.2	6.0 ± 2.5
>1	1.4 ± 0.7 ^{ab}	1.9 ± 0.9 ^{ab}	2.2 ± 0.1	3.7 ± 0.2	1.0 ± 0.6	6.2 ± 2.8	0	0
F value	4.83	17.90	0.33	0.78	0.41	1.34	0.43	1.61
P	0.008	< 0.001	0.72	0.46	0.52	0.26	0.51	0.21

Larinus: 0 insects n = 490; 1 insect n = 242; >1 insect n = 8.
Urophora affinis: 0 insects n = 82; 1 insect n = 103; >1 insect n = 555.
Urophora quadrifasciata: 0 insects n = 714; 1 insect n = 21; >1 insect n = 5.
Metzneria paucipunctella: 0 insects n = 735; 1 insect n = 5; >1 insect n = 0.
Pairwise comparisons made with the Tukey Honest Significant Difference (HSD) method; p < 0.05.

Table 3. ANOVA with Tukey multiple comparisons of means to detect geographic variation in seed bank and insect density. Results indicate the mean count of total seed was greater at the Road site, while the % viable seed count was greater at the Hill site. Mean density of *Larinus minutus* was lower at the Road site while mean density of *Urophora affinis* was lower at the Hill site.

Geographical variation of spotted knapweed seed bank and insect numbers per seed head (mean ± SEM).

	Seed bank (n = 57)		Insect numbers per seed head (n = 740)			
Site	Seed bank (total seeds/m2)	Seed bank (% viable seeds)*	*Larinus minutus*	*Metzneria paucipunctella*	*Urophora affinis*	*Urophora quadrifasciata*
blue	1711.1 ± 355.5 ^{a}	1.7 ± 1.0 ^{a}	0.37 ± 0.03 ^{a}	0.01 ± 0.01	3.19 ± 0.1 ^{a}	0.06 ± 0.02
road	4226.0 ± 734.9 ^{b}	2.7 ± 0.6 ^{a}	0.24 ± 0.03 ^{b}	0.00 ± 0.00	3.36 ± 0.1 ^{a}	0.06 ± 0.02
hill	1547.4 ± 288.9 ^{a}	20.5 ± 3.4 ^{b}	0.43 ± 0.03 ^{a}	0.01 ± 0.01	2.36 ± 0.1 ^{b}	0.02 ± 0.01
F value	8.714	26.210	9.998	0.167	19.330	1.710
P	< 0.001	< 0.001	< 0.001	0.846	< 0.001	0.182

Seed bank n = 57 (blue n = 18; road n = 20; hill n = 19).
Insect numbers n = 740 (blue n = 253; road n = 237; hill n = 250).
*Ruptured and germinated seeds as a percentage of the total seed bank.
Pairwise comparisons made with the Tukey Honest Significant Difference (HSD) method; p < 0.05.

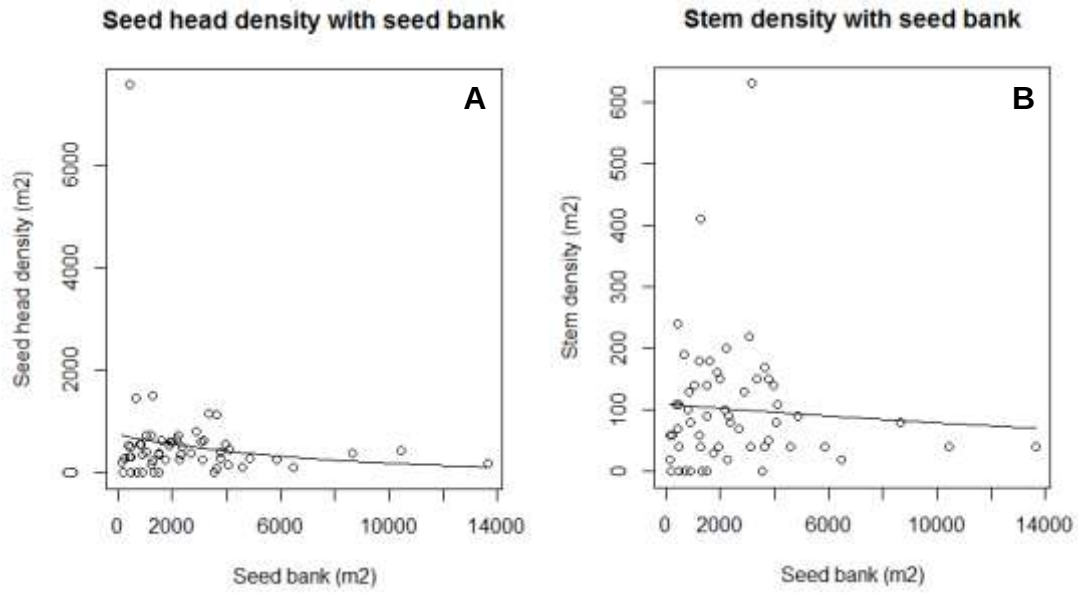
Table 4. ANOVA with Tukey multiple comparisons of means to detect geographic variation in viable seed count per seed head and vegetation parameters. Results suggest that at the Blue site, means of viable (black) seeds were lower and non-viable (other) seeds were higher; and the density of rosettes and seedlings were higher compared to the other sites

Geographical variation in reproductive potential and vegetation parameters of spotted knapweed (mean ± SEM).

	Seeds per seed head (n = 740)		Vegetation parameteres (n = 57)				
Site	Seeds (black)/head	Seeds (other)/head	Rosettes/m2	Seedlings/m2	Seed heads/m^2	Bolted plants/m2	Stems/m2
blue	1.8 ± 0.2 ^{a}	5.4 ± 0.3 ^{a}	7.28 ± 1.5 ^{a}	6.28 ± 2.6 ^{a}	36.9 ± 7.9	3.78 ± 0.7	8.89 ± 1.5
road	2.6 ± 0.3 ^{b}	3.0 ± 0.2 ^{b}	4.50 ± 0.7 ^{ab}	4.30 ± 1.6 ^{ab}	39.6 ± 7.3	3.15 ± 0.5	8.75 ± 3.1
hill	2.2 ± 0.2 ^{ab}	2.3 ± 0.2 ^{b}	3.63 ± 0.8 ^{b}	0.32 ± 0.2 ^{b}	87.9 ± 37.9	3.95 ± 0.6	12.84 ± 2.2
F value	3.96	40.68	3.437	3.086	1.59	0.541	0.951
P	0.0194	< 0.001	0.039	0.054	0.213	0.585	0.393

Seed head n = 740 (blue n = 253; road n = 237; hill n = 250).
Vegetation parameters n = 57 (blue n = 18; road n = 20; hill n = 19).
Pairwise comparisons made with the Tukey Honest Significant Difference (HSD) method; p < 0.05.

Figure 2. Generalized linear models with *poisson* distribution; seed head density (R^2 = 0.05) and stem density (R^2 = 0.01) against seed bank. While the coefficients for both models are significant ($p < 0.001$) the residuals compared to the degrees of freedom indicate severe overdispersion in the data, possibly due to the outlying data points (confirmed with the `outlier()` function from the 'outliers' package). Transformation of the data did not produce a near normal distribution that was suitable for linear or generalized linear models; logistic regression was not used because the data were not binary.



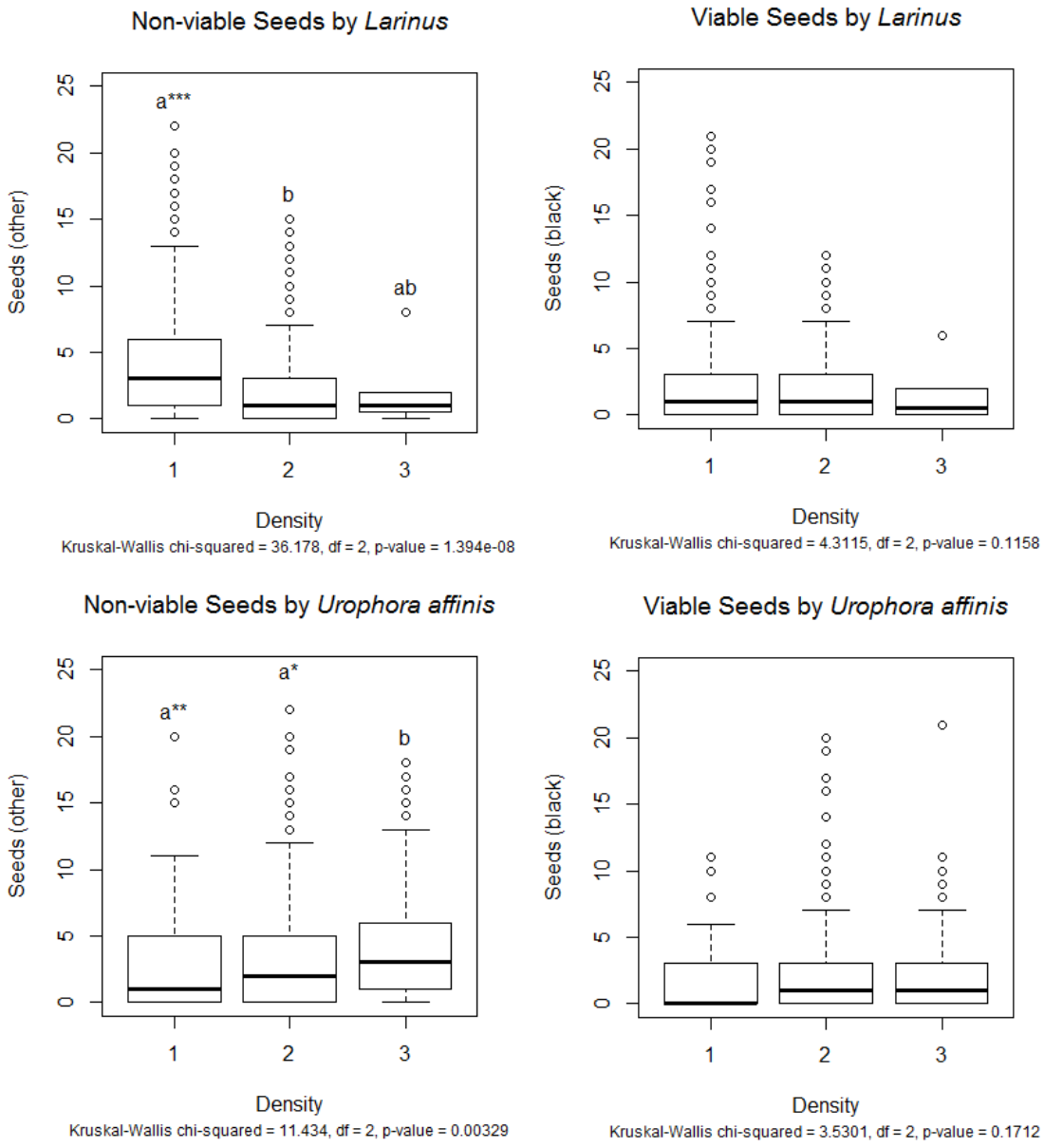


Figure 3. Kruskal-Wallis rank sum tests with pairwise comparisons followed by Conover's test for multiple comparisons of independent samples using the Bonferroni *p*-value adjustment method. Seed number and insect density based on the count per seed head sample. Density (counts) of each insect is presented in 3 categories (approximately tertiles). Categories are as follows: *Larinus minutus*: 0=1 (*n*=490), 1=2 (*n*=242), and 2=3 (*n*=8); *Urophora affinis*: 0=1 (*n*=82), 1-4=2 (*n*=489), and 4-10=3 (*n*=169). *Chaetorellia acrolophi* was not recorded in the seedhead samples. In the 740 samples (individual seedheads), *Metzneria paucipunctella* was present in 5 seedheads and *Urophora quadrifasciata* was present in 26 seedheads. Density of *Larinus minutus* and *Urophora affinis* suggest lower and higher means (respectively) of non-viable seeds. No difference in means of either "black" or "other" seed count was detected when comparing seedheads with versus without *Metzneria paucipunctella* or *U. quadrifasciata*. Letters [a, b] indicate significant differences <0.05*; <0.01**; and <0.001***.

Summary

Approximately 95% of the seed heads sampled contained at least one of the biological control insect species. *Urophora affinins* was the most abundant of the five species and *Larinus minutus* was the second most abundant species while *Chaetorellia acrolophi* was not recorded in any of the seed head samples. There appears to be a negative relationship between *Urophora* spp. and between *Larinus minutus* and *Urophora affinins*.

ANOVA results found differences in both viable and non-viable seed head seeds infested with *Larinus minutus*. However, Kruskall-Wallis tests showed an increase in non-viable seeds counted in seed heads explained by the density of both *Larinus minutus* and *Urophora affinins* while there was no significant relationship detected between either of these two species and the number of viable seed remaining in each seed head. Interesting trends were evident with geographic variation: where *Urophora affinins* density was lowest (Hill site), the % viable seed in the seed bank was significantly higher; and where density of *Larinus* was lowest (Road site), there was an increase in the means (total seeds) of the seed bank.

It was intended for this analysis to be used for comparison with previous studies (Story et al. 2008), however the consistency in methods is unconfirmed and the variability within and between these studies would require additional investigation. For future analysis I recommend non-parametric methods including ordination techniques such as nonmetric multidimensional scaling accompanied by permutational ANOVA as offered by the `adonis()` function in R's 'vegan' package.

References

Callaway, R. M.; DeLuca, T. H.; Belliveau, W. M. (1999). "Biological-Control Herbivores May Increase Competitive Ability of the Noxious Weed *Centaurea maculosa*". Ecology. 80 (4): 1196–1201.

DiTomasso, J.M. (2000). "Invasive weeds in rangelands: Species, impacts, and management". Weed Management. 48 (2): 255–265.

Hirsch, S. A.; J. A. Leitch (1996). ". The affect of knapweed on Montana's economy". Agricultural Economics.

John R. Lacey; Clayton B. Marlow; John R. Lane (1989). "Influence of spotted knapweed (*centaurea maculosa*) on surface runoff and sediment yield". Weed Technology. 3 (4): 627–631.

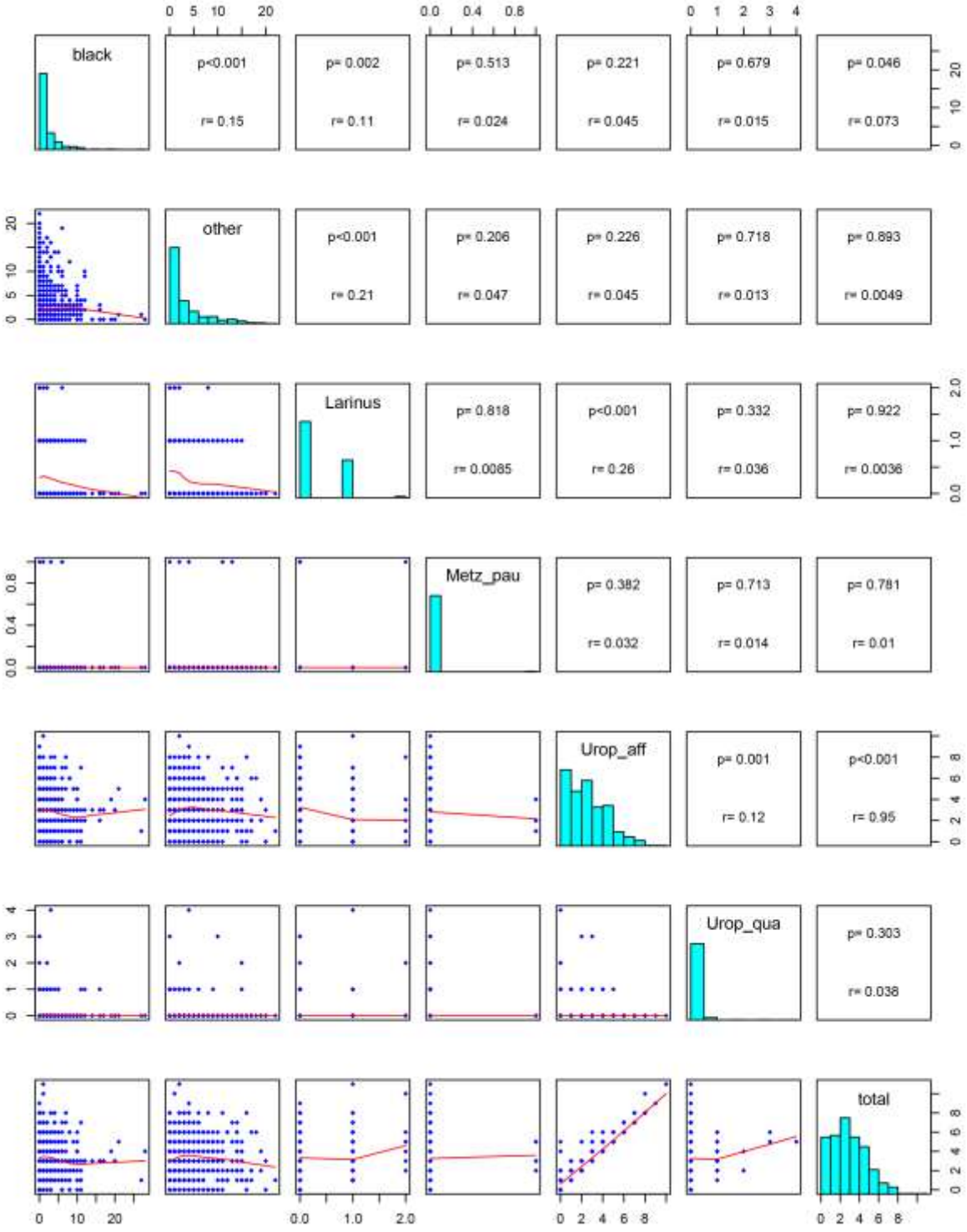
Mueller-Scharer, H.A.; Schroeder, D. (1993). "The biological control of Centaurea spp. in North America: do insects solve the problem?". Pest. Sci. 37: 343–353.

Roger L. Sheley; James S. Jacobs; Michael F. Carpinelli (April–June 1998). "Distribution, Biology, and Management of Diffuse Knapweed (*Centaurea diffusa*) and Spotted Knapweed (*Centaurea maculosa*)". Weed Technology. 12 (2): 353–362.

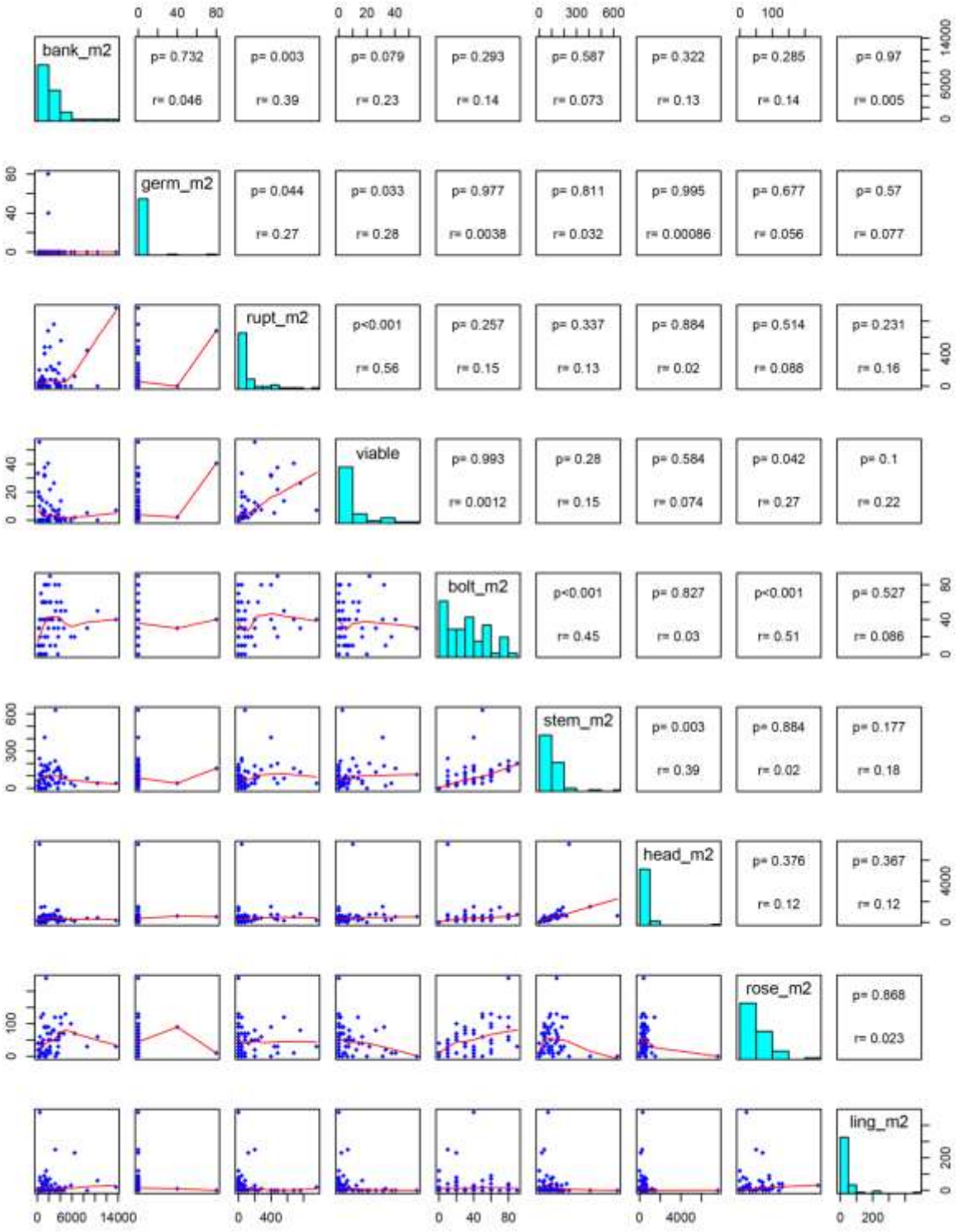
Story, Jim M., et al. (2008) "Influence of seed head–attacking biological control agents on spotted knapweed reproductive potential in western Montana over a 30-year period." Environmental Entomology 37.2: 510-519.

Team RDC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Version 3.3.1. 2016.

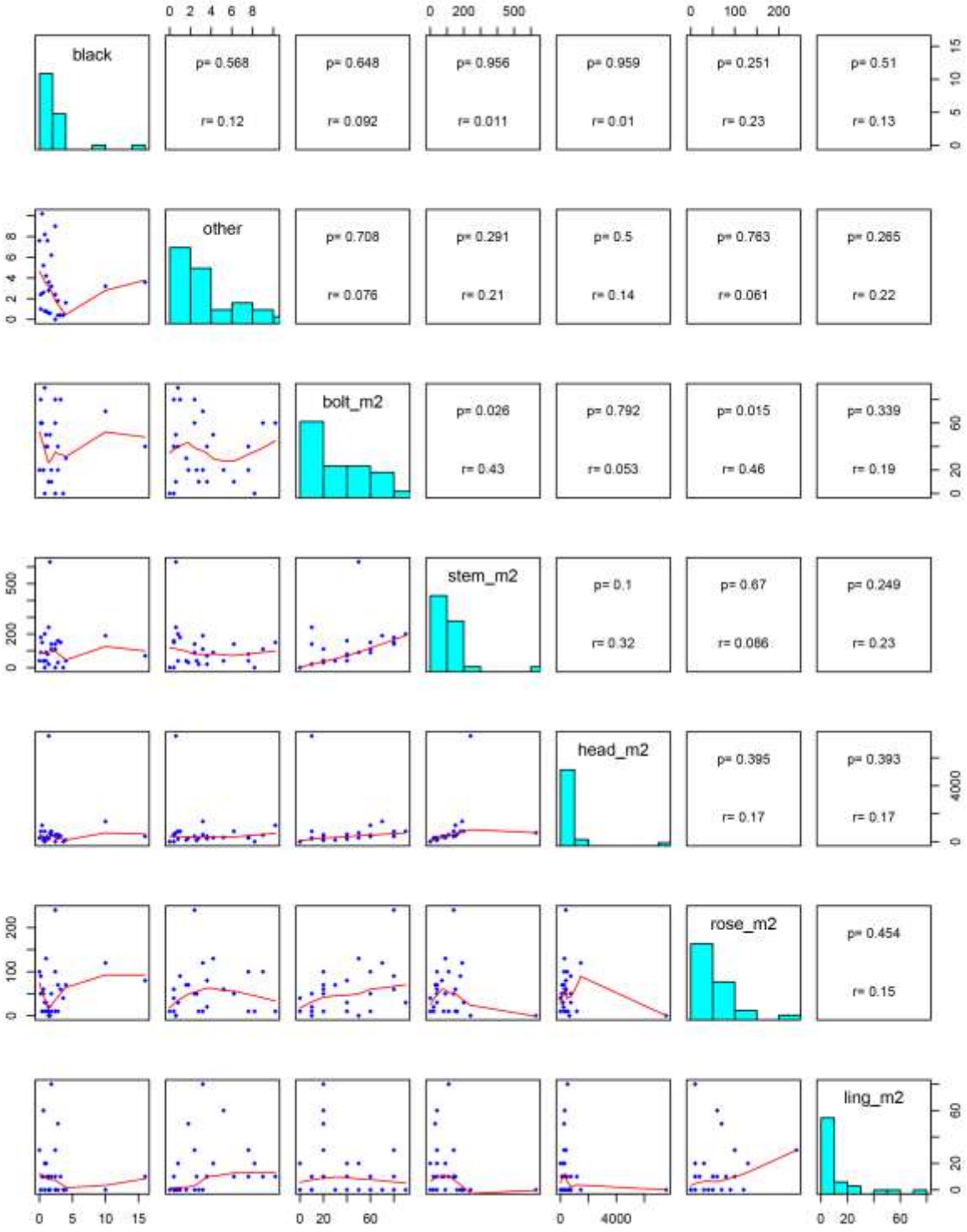
Appendix A: Correlations – Seed head seeds and seed head insects (n = 740).



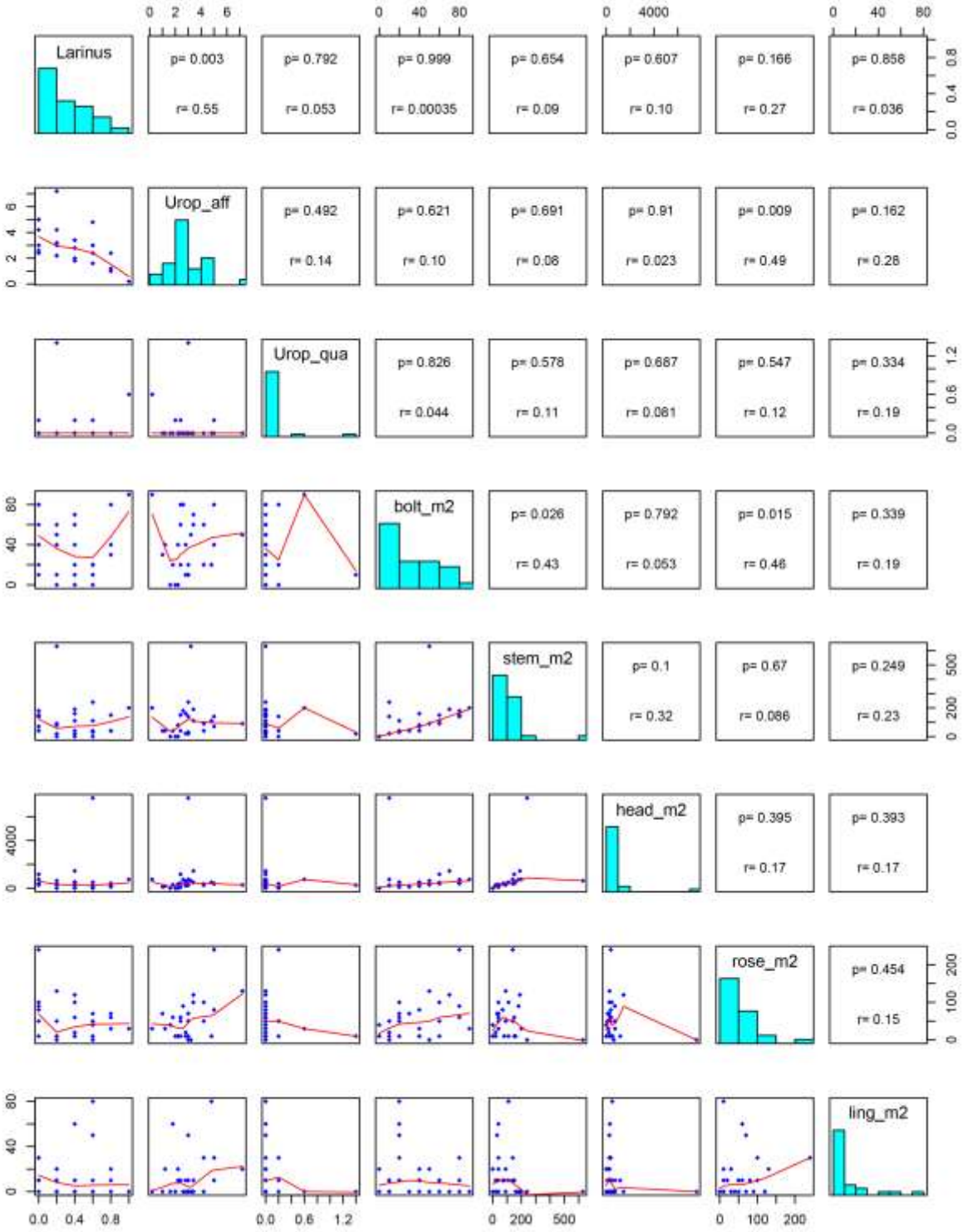
Appendix A: Correlations – Seed bank and vegetation parameters (n = 57).



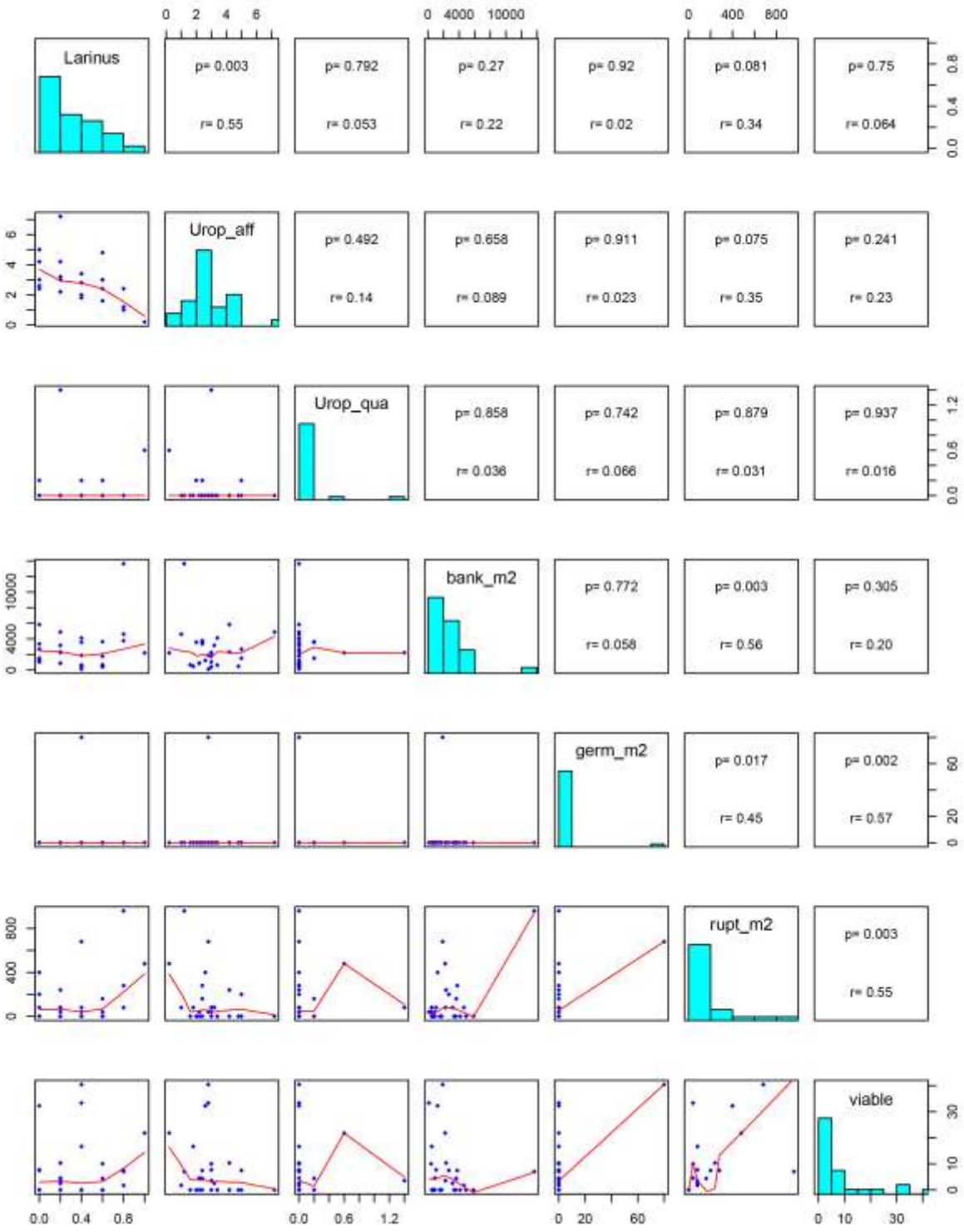
Appendix A: Correlations – Seed head seeds and vegetation parameters (n = 27).



Appendix A: Correlations – Seed head insects and vegetation parameters (n = 27).



Appendix A: Correlations – Seed head insects and seed bank (n = 27).



Appendix A: Correlations – Seed head seeds and seed bank (*n* = 27).