*\* 2013 transplant data & 2013 -> 2014 survivors not discussed here.*

**TRANSPLANT DATA ANALYSIS WORKFLOW**

**R-code files:**

00\_SetDirectories.R – sets local directory files to computer.

01\_DataStruc\_check.R – visualization & quality check of data & data structure.

02\_Vital\_rate\_regression.R – functional form for size specific vital rate regression equations.

03\_Major\_Source\_of\_variation.R – quick visualization of major source of spatial & temporal variability in vital rates.

04\_Plot\_level\_variables\_VR.R– relationship between vital rates and plot-level environmental variables.

05\_PCA\_plot\_microsite\_variables – relationship between vital rates and plot-level environmental variables.

**Selecting size-specific functional forms for vital rate regression functions**:

R-code: 02\_Vital\_rate\_regression.R

*Temporary equations for the basic structure of random effects models are listed at the bottom. Vital rate regression functions & associated environmental coefficients will be later fit individually for each site, but the purpose of this to justify the functional relationship of each vital rate with size (linear, quadratic, cubic or just a flat line (intercept only)).*

From the available data, it is only possible to create within season vital rate regressions for size specific survivorship, growth, probability of flowering and fecundity, rather than estimates of λ (until the final July 2015 census).

Workflow for model selection follows the Merow *et al* (2014) Tutorial: [Advancing population ecology with integral projection models: Appendices](http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12146/suppinfo).

Following Chapter 4, Appendix D: **Vital Rate Regressions for Integral Projection Models**.

*Assuming here that overall model form for size specific functions should be consistent across sites, but differ in relation to environmental variables.* Models are developed from a global dataset with two hierarchal nested random effects for site and plot (1|site/plot).

Possible model complexity for each vital rate with size ranges from a basic intercept, to a linear relationship with size, then to size2 and size3 just to see if a complex relationship is nessesary. Support for a more complex model is supported by AIC scores and visually justified by successively subsampling the data at the plot level.

Full model: All plants within a plot are used for the model.

75%: 25% of plants within each plot are excluded & models run.

50%: 50% of plants within each plot are excluded & models run.

25%: 75% of plants within each plot are excluded & models run.

**Main components of analysis:**

1. **Exploratory analysis of plot-level environmental variables**.
   1. R-code: 03\_Major\_Source\_of\_variation.R

Which spatial scale (within sites vs. between sites) explained most of the variation for each of the four vital rate regressions (survivorship, growth, pflower and fecundity)?

*Rational:* *Simple visualization and justification for further exploration into plot level variables.*

* + 1. Include random variable in global model: (1|site/plot) for each vital rate. Produce barplot of % variance explained by across sites & within sites against each vital rate for spring growth and summer growth.

(Recreate plot similar to [Fig 2 in Ramula 2014](http://link.springer.com/article/10.1007%2Fs00442-013-2869-3))

* 1. R-code 04\_Plot\_level\_variables\_VR.R

Regardless of site-level climatic differences - were vital rates strongly dependent on any of the measured plot-level variables? Were the functional forms of these relationships also similar across sites?

*Rational: If the effects of site-level climatic differences on vital rates were subtle in comparison to plot level differences, then it will be necessary to incorporate these plot-level variables into the final models.*

Approach: Run exploratory regressions for each site, for the four vital rates functions against several of the key plot-level variables. In addition to measured variables, also run regressions on PCA 1, PCA 2 & PCA 3 (from the plot-level environmental matrix).

(Recreate applicable part of [Fig 3 & 4, in Diaz et al 2014](http://onlinelibrary.wiley.com/doi/10.1111/1365-2745.12215/abstract) with transplant data)

*This will give a better idea of the dependency of the vital rates to the plot level environmental variables as well as whether or not they should be incorporated in final analysis.*

* 1. 05\_PCA\_plot\_microsite\_variables

From the plot-level environmental variables collected:

* + 1. Were any sites significantly different from each other? (run perMANOVA or MRPP on plots with site as a factor)
    2. Were the selected transplant plots similar to environmental variables collected from wild *M. cardinalis*? (use MRPP because design heavily unbalanced; wild individual measurements (~ 20 – 30/creek) nested within specific creeks (Coast, Rock, Calapooya, Canton)). *Need to revisit details of this analysis.*

*If only a very small number of plots are extreme environmental outliers then it may be easier to exclude them rather than trying to incorporate microsite variables into final analysis.*

1. **RANGE LIMIT AND SITE-TYPE DIFFERENCES:**

*Did the vital rates of transplants differ across sites within and beyond the range limit or across occupied vs. unoccupied sites?*

* 1. Did vital rates differ across sites located within & beyond the range limit?
     1. Method 1: For each vital rate, use a hierarchal mixed effects model with sites and plots as random effects nested within the site-type regional group i.e. (1|siteType/site/plot).
        1. REML used first to compare across different nested models structures (is it possible to drop random terms of plot or site?).
        2. Then REML set to false & ML used to compare the inclusion & exclusion of different fixed effects.
     2. Method 2: Make site level estimates and compare values across sites from site level confidence intervals: Details for generating site level confidence intervals from mixed effects models still needs to be worked out.
  2. Repeat the same process as above for occupied vs. unoccupied sites as the ‘siteType’ variable rather than within or beyond the range limit.
* Exclude all sites beyond the range for this test.

1. **RELATIONSHIP BETWEEN VITAL RATE EQUATIONS AND SDMS PREDICTIONS**:
   1. Regressions between site-level vital rate estimates and raw suitability output from SDM model (occupancy paper & this study should use same models: both focus on niche).
      1. SDM models to compare:
         1. Basic ensemble output and individual models used in occupancy paper.
         2. Simplified SDM models, potentially with varying levels of complexity and # of variables e.g.
            * Model 1: Simple: few variables (less than 4);
            * Model 2: Medium complexity; 4-8 variable
            * Model 3: Complex: many variables (8)

**\*** *Will hopefully be able to produce this series of models varying in their levels complexity as a sensitivity test for Occupancy Paper* – from the list of revisions*, then use those same models here.  
(Get two bird with one stone here?)*

* 1. Regressions between site-level vital rate estimates and other climatic variables (from ClimateWNA and direct field temperatures from HOBO loggers – e.g. min/max daily temperatures in June)

*The purpose of these multiple comparisons between vital rates, climate variables and SDM output scores is exploratory to gain an understanding & diagnosis of why there are inconsistencies between transplant data and SDM predictions, but numerous comparisons already concerning.*

*[4 vital rates \* ~ 7 climate variables] + [4 VR \* (6 SDM models \* ~ 3 levels of complexity)] = ~ 105 climate-vital rate comparisons from exploratory analysis.*

**Vital rates of Interest:** Most basic (non-intercept) model run for each site (R package lme4)

**Survivorship** – Survivorship from start to end of time period (distribution: binomial)

**Growth** – Growth (given survivorship), from start to end of time period (distribution: Gaussian/normal)**.**

**Flower** – Size specific probability of flowering (distribution: binomial)

**Fecundity** - size-specific fecundity (distribution: Poisson)

**\*** In lmer4, when trying to look across site-type, unable to use Poisson distribution with any nested random effect. For some reason lme4 & nlme gives an error message with the Poisson distribution when there is more than one random effect or when random effects are nested. Considering log transformation & then checking residual plot for normal distribution.

\* If the design using random effect is unbalanced (unequal number of plants within plots); F-statistics are not F-distributed and standard ANOVA tables don’t work. But REML can handle unequal sample size so OK?

**Data structure:**

Spatial structure:

**Site-type** (*Within range* OR *beyond range*) (*Occupied* or *Unoccupied*)

**Site** (*Rock, Coast, Mosby*… ect.

**Plots** (P01, P02 ….. ect. \*with unique ID

**Individuals** (*Individual plants*)

Monitoring periods:

For exploratory analysis:

1. MAY (Planting) **- >** JUNE (Post transplant survivorship census).
2. JUNE **->** JULY CENSUS (spring growth).
3. JULY CENSUS **->** SEPT CENSUS (summer growth).

For final analysis:

Post-transplant survivorship -> SEPT CENSUS

**OTHER NOTES ON MIXED EFFECT MODELS:**

* Should a random effect in a nested design be dropped if it is deemed non-significant: probably not, would be considered sacrificial pseudo-replication.

***Ozgul 2010 Nature***

* *IPMs for later >2000 & before 2000, year treated as a random factor. In my data within & beyond could be compared with site treated as a random factor.*
* *Anova with the reduced model to test signifgance of terms via a likelihood ratio test*