Niche-Based Distribution Modelling of Koala Habitat

Allen McIlwee

Senior Scientist GIS and Spatial Analysis

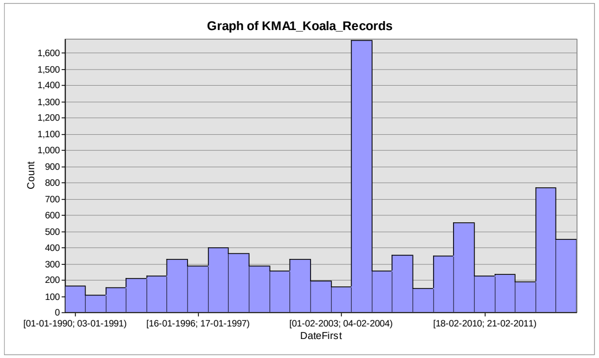
Office of Environment and Heritage

- **Purpose**
  - Map the potential distribution of Koalas across Northern Rivers CMA in terms of its probability of occurrence
  - Understand the important environmental settings and relationships that underpin potential habitat

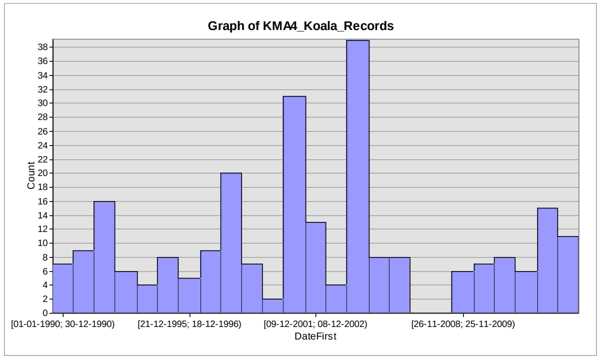
Note: it's the distribution of potentially suitable habitat (not the actual distribution of koalas) that is modelled. Not all suitable habitat will be occupied.

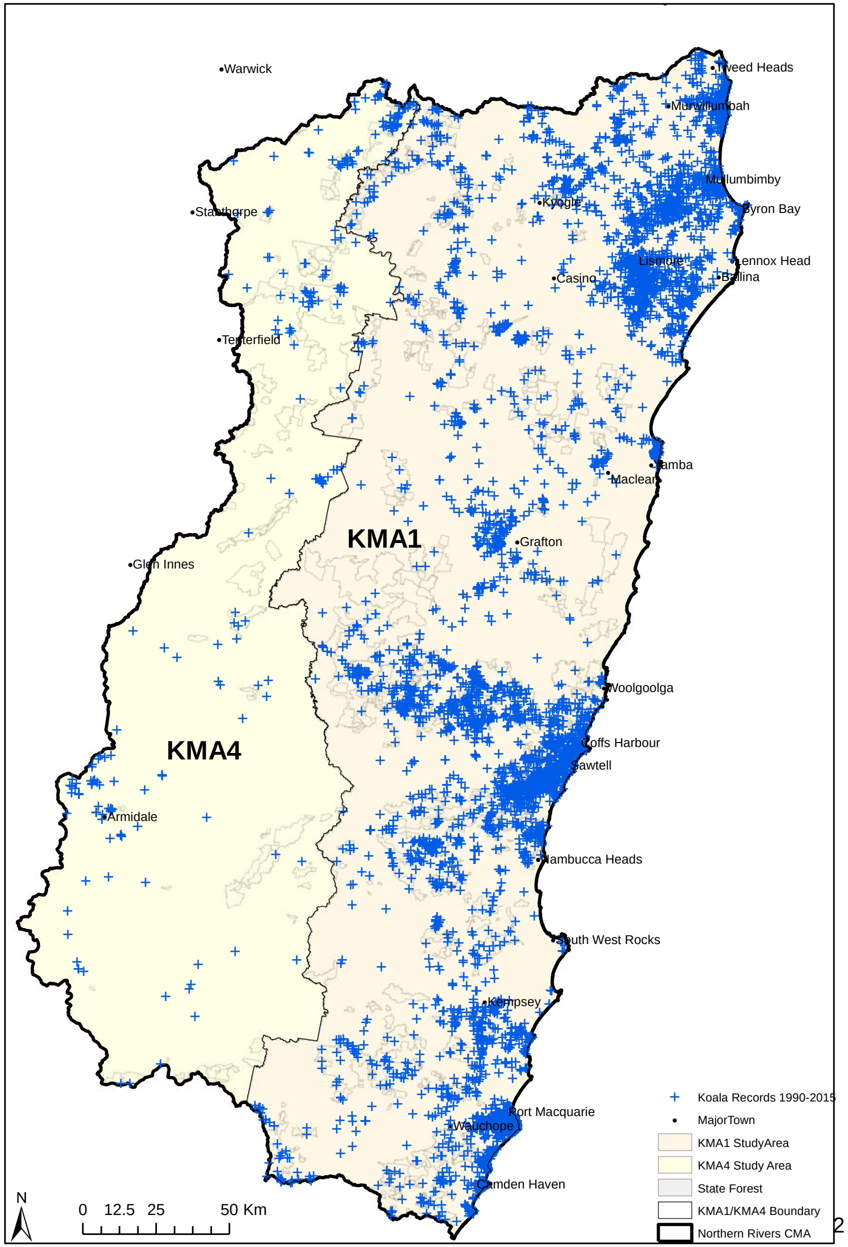
Koala records

- 9083 Koala records across CMA
  - **KMA1**
    - 8822 (97.2%) records



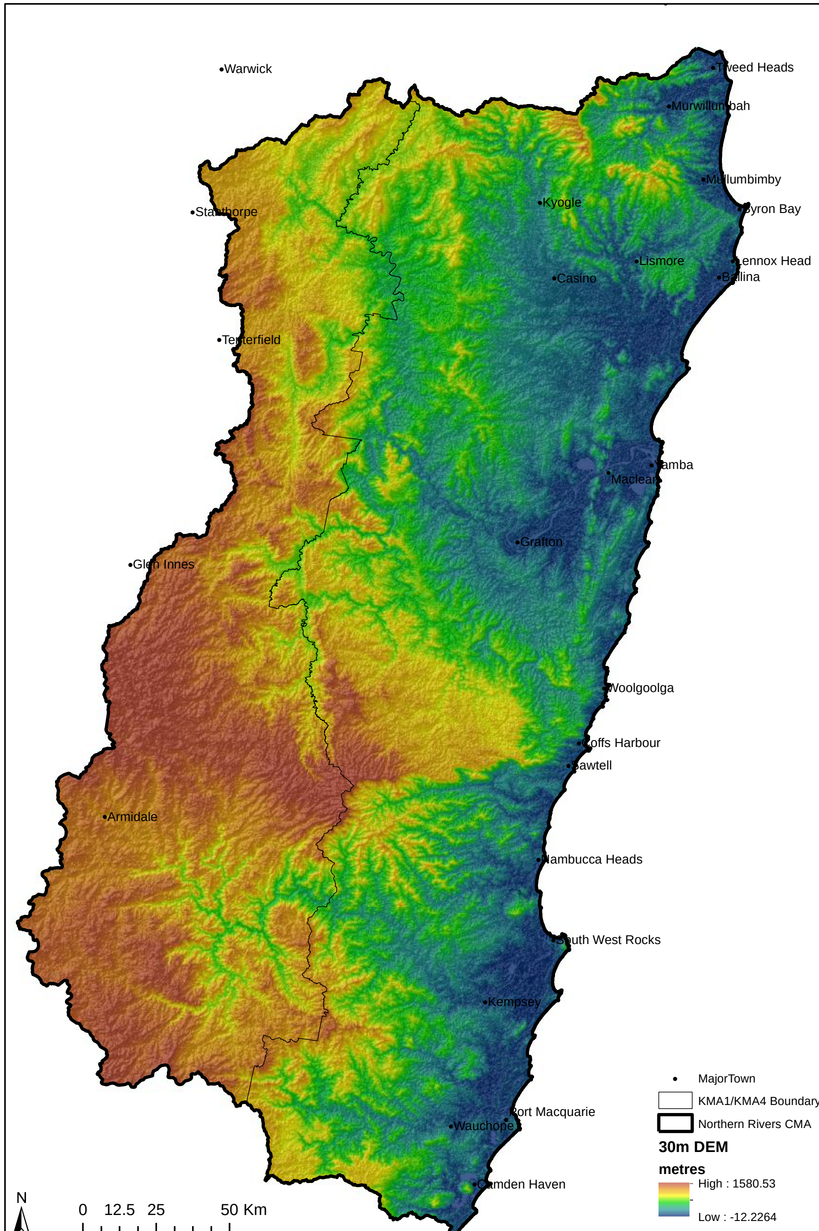
  - **KMA4**
    - 261 (0.28%) records





Predictor variables

- 142 Possible predictors captured at 30m pixel level across CMA
- Types of predictors
  - Distance based metrics (n=5)
  - Landform and terrain (n=19)
  - Climate – radiation and energy (n=8)
  - Climate – temperature (n=17)
  - Climate – rainfall (n=17)
  - CSIRO Soil Profile (n=49)
  - Soil Minerals (n=6)
  - Soil NIR spectra (n=6)
  - Radiometrics (n=4)
  - Geophysics (n=10)
  - Remote Sensing – point in time mosaics (n=4)
  - Remote Sensing – 24yr within season temporal stats (n=24)



Selection of absent sites

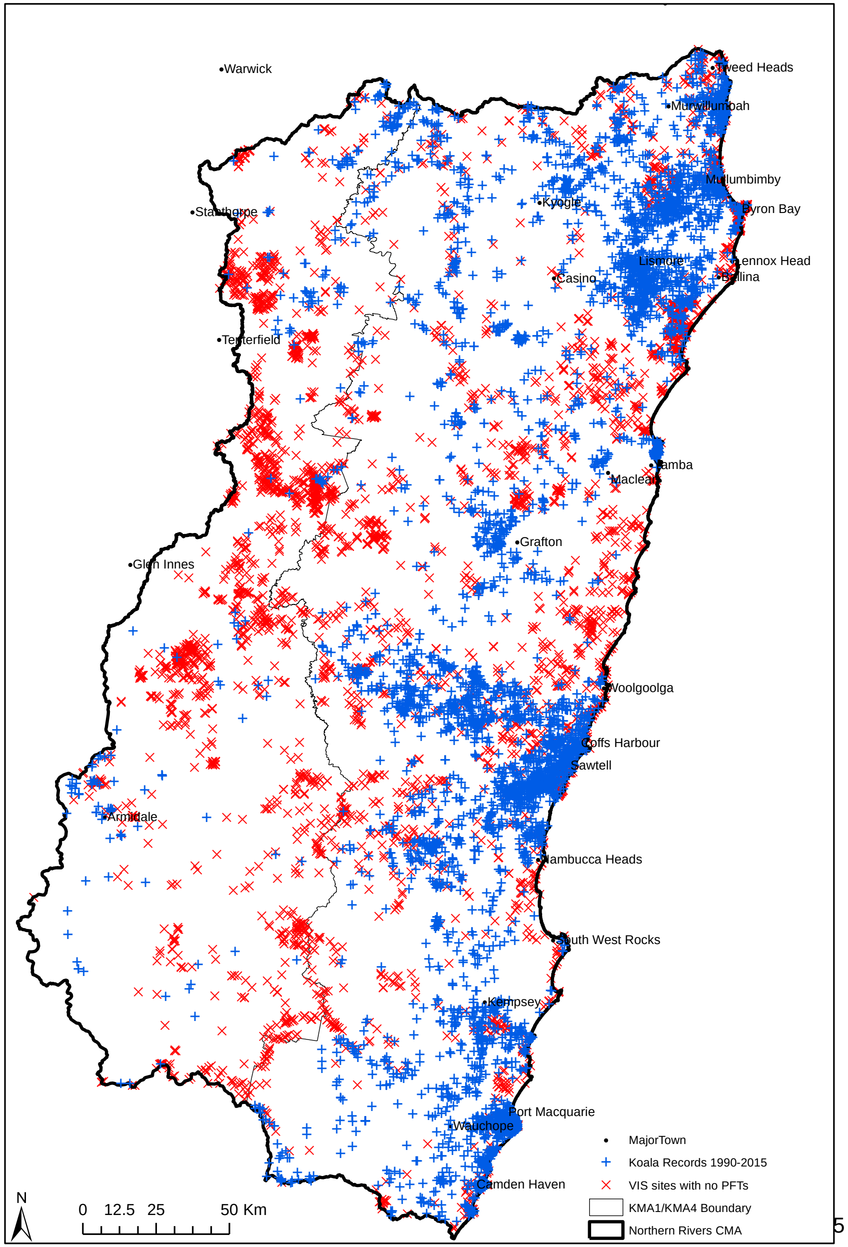
- 6881 Full floristic VIS plots across CMA
- **KMA1**
  - 4377 VIS plots
  - 2098 (48%) PFTs present
  - 2279 (52%) PFTs absent
- **KMA4**
  - 2504 VIS plots
  - 1027 (41%) PTFs present
  - 1477 (59%) PTFs absent

List of Primary Food Trees (PFTs)

Common Name	Scientific Name	KMA1	KMA4
White Box	*E. albens*		*
Orange Gum	*E. bancrofti*	*	*
Grey Gum	*E. biturbinata*	*	*
Blakely's Red Gum	*E. blakelyi*		*
Large-fruited Grey Gum	*E. canaliculata*	-	
Monkey Gum	*E. cypellocarpa*		-
Tumbledown Red Gum	*E. dealbata*		*
Dwyers Red Gum	*E. dwyeri*		-
Slaty Red Gum	*E. glaucina*	*	*
Silver-top Stringy	*E. laevopinea*		*
Yellow Box	*E. melliodora*	*	**
Tallowwood	*E. microcorys*	*	*
Grey Box	*E. moluccana*	*	**
Parramatta Red Gum	*E. parramattensis*	-	
Small-fruited Grey Gum	*E. propinqua*	*	*
Swamp Mahogany	*E. robusta*	*	*
Forest Red Gum	*E. tereticornis*	*	**
Ribbon Gum	*E. viminalis*		*

Presence – absence dataset

- **Basis for models**
  - Koala presence–absence data have not been sampled in a systematic way
  - Here absences reflect a genuine lack of Primary Food Trees (but not koalas)
  - Absences are used to constrain model on the premise that koalas are largely absent from areas with no PFTs.
  - Models can tolerant a margin of error in relation to false absences



Modelling Approach

- **Boosted Regression Trees (BRT)**
  - Ensemble method for fitting statistical models
  - Fundamentally differs from conventional modelling techniques that aim to fit a single parsimonious model (e.g. GLM)
  - As name implies - uses Regression Trees and Boosting
- **Regression Trees**
  - Set of hierarchical non-parametric decision rules that each terminate in a final decision (e.g. the predicted presence or absence of koalas)
  - In BRTs thousands of Regression Tree models are combined
- **Boosting**
  - Technique for improving model accuracy. Based on idea that it is easier to find and average many rough rules of thumb, than it is to find a single, highly accurate prediction rule

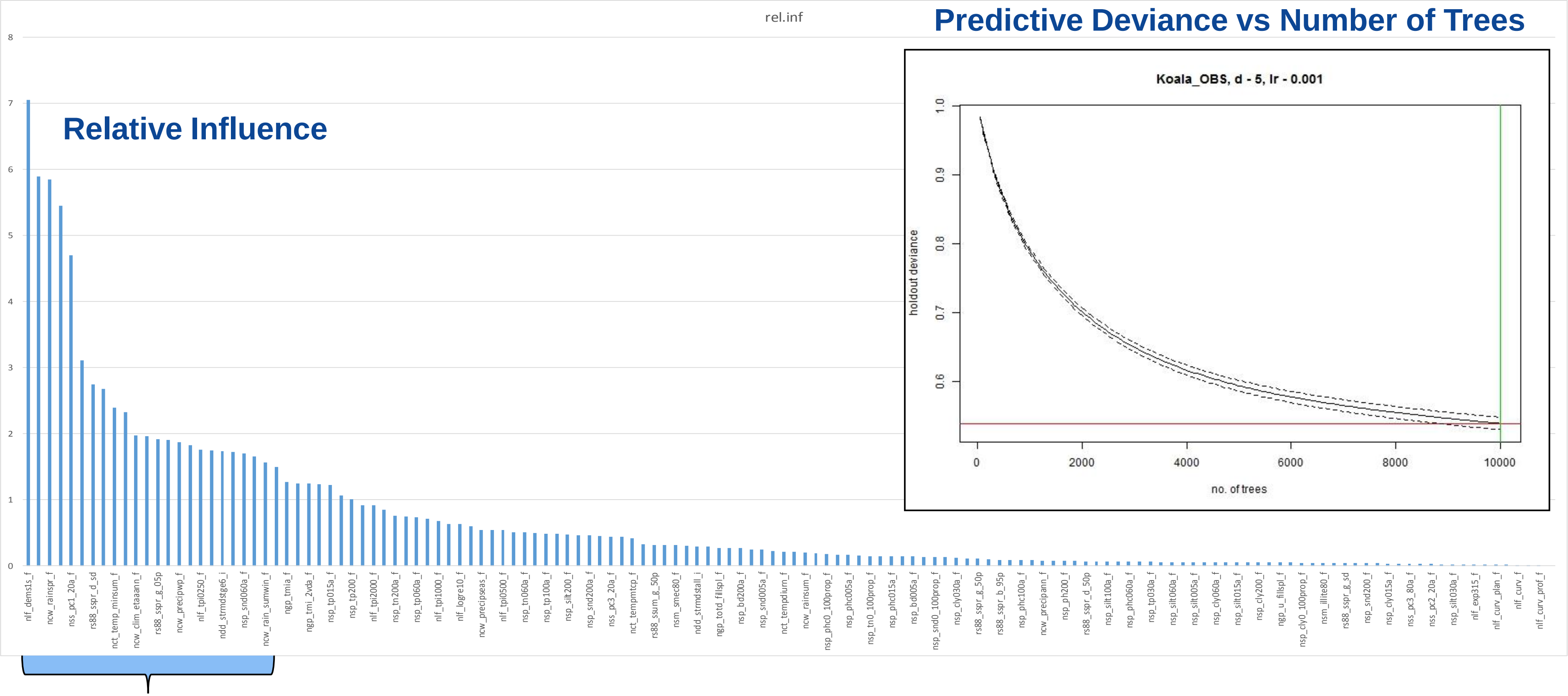
Modelling Approach

- **Boosted Regression Trees (BRT)**
  - Final BRT model is a linear combination of thousands of trees, where random subsets (i.e. 75%) of data are used to fit each new tree
  - Essentially it's a regression model where each term is a tree.

- **Why use BRTs?**
  - Most reliable of 10 methods tested on EEC communities
  - Can handle NA values in predictor variables
  - Scaling & normalization is unnecessary
  - Can handle multiple highly correlated variables
  - Can models different types of loss functions.
    - Bernoulli: logistic model for presence absence data
    - Multinomial: where more than one class (produces probability matrix)

Modelling Approach

- ## Modelling Process
  - Run gbm.step model with all predictors to identify initial subset based on relative influence values
  - Learning rate (*lr*) set very low (i.e. 0.001) with a bagging fraction of 75% and tree complexity of 5.

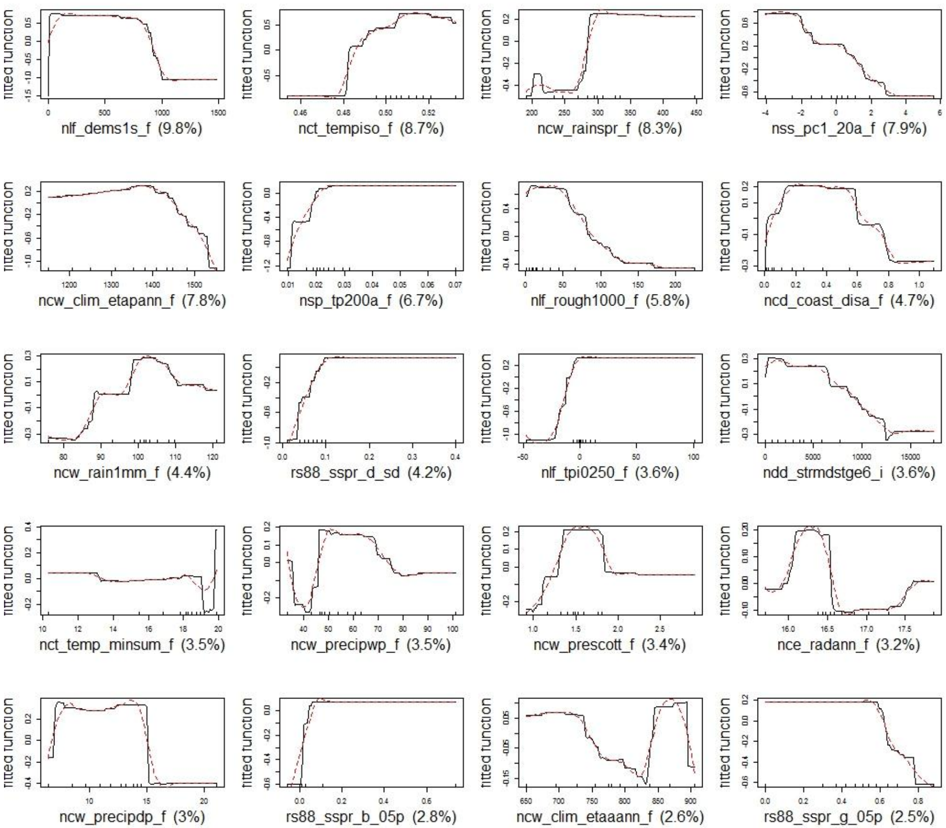


Initial subset of predictors

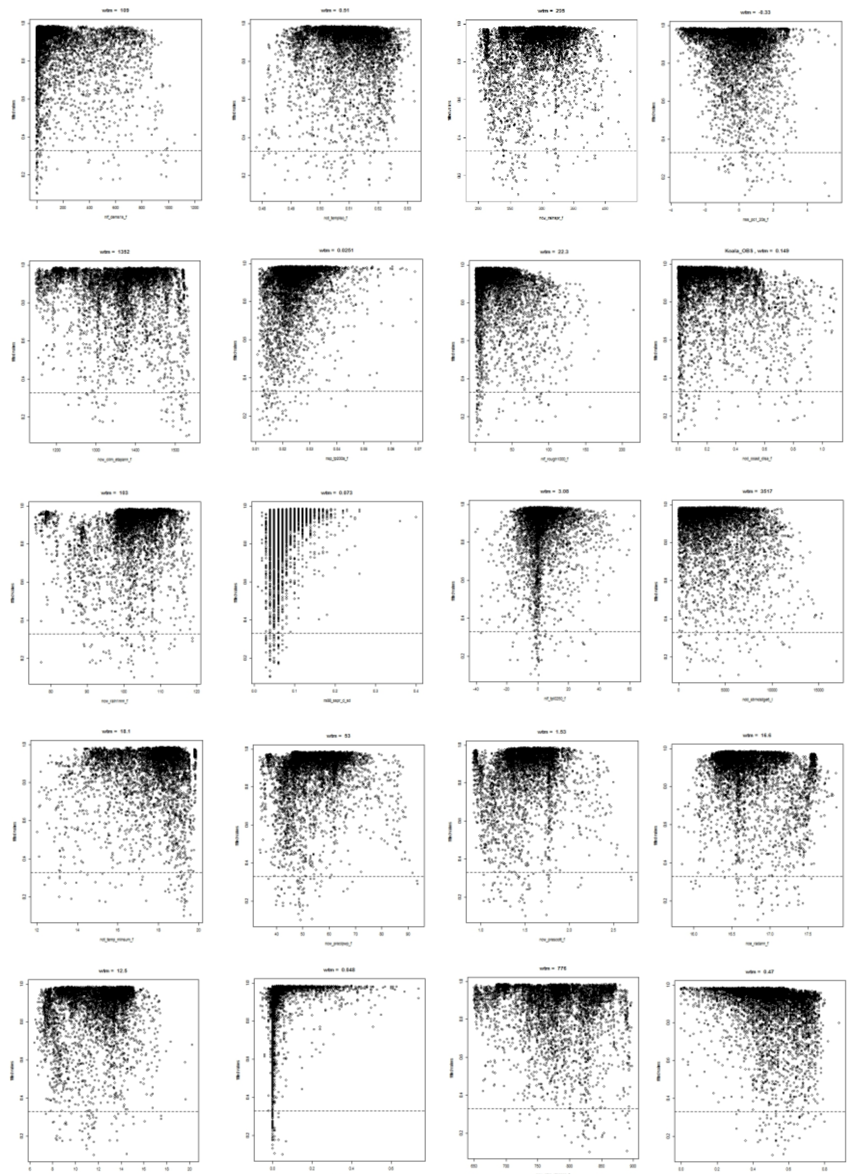
Modelling Approach

- Rerun gbm.step with set of predictors
- Plot fitted functions & fitted values to check whether predictors & modelled relationships make ecological sense

Fitted functions



Fitted values (presence only)

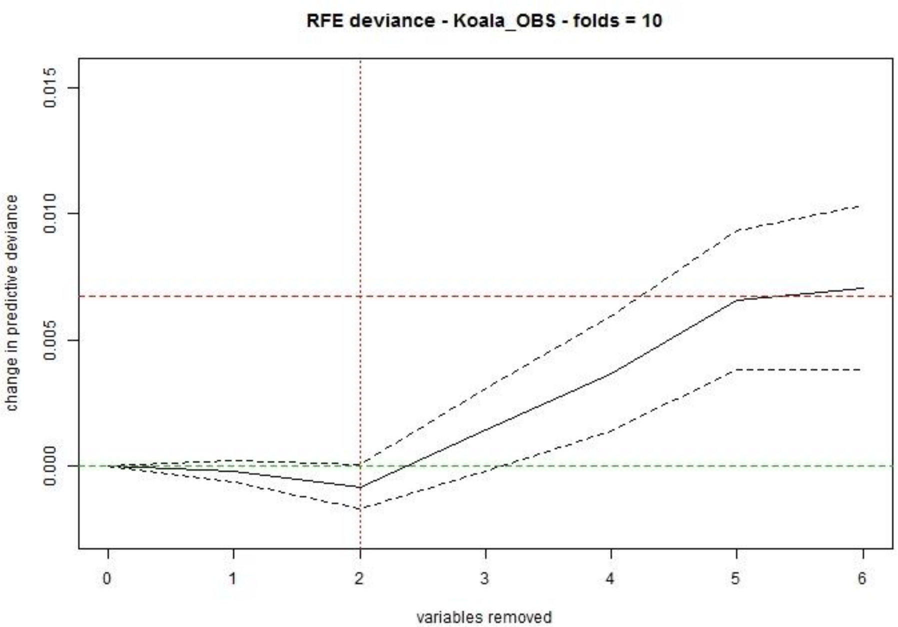


Office of Environment & Heritage
NSW GOVERNMENT

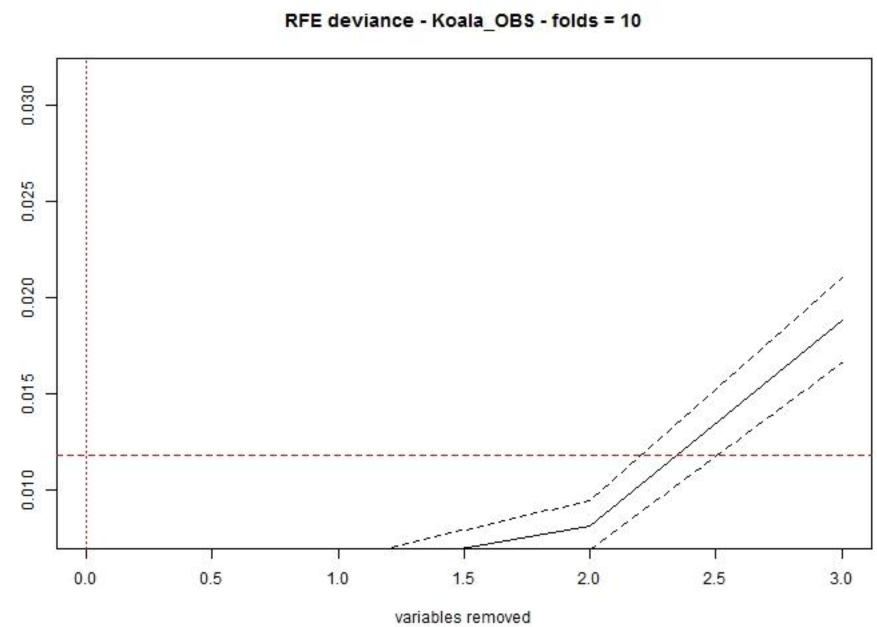
Modelling Approach

- **Final step**
  - Use gbm.simplify algorithm to test for and remove redundant variables
  - This takes an initial cross-validated model produced by gbm.step and assesses the potential to remove predictors using k-fold cross validation
  - The function returns a list containing the mean change in deviance and its standard error as a function of the number of variables removed.

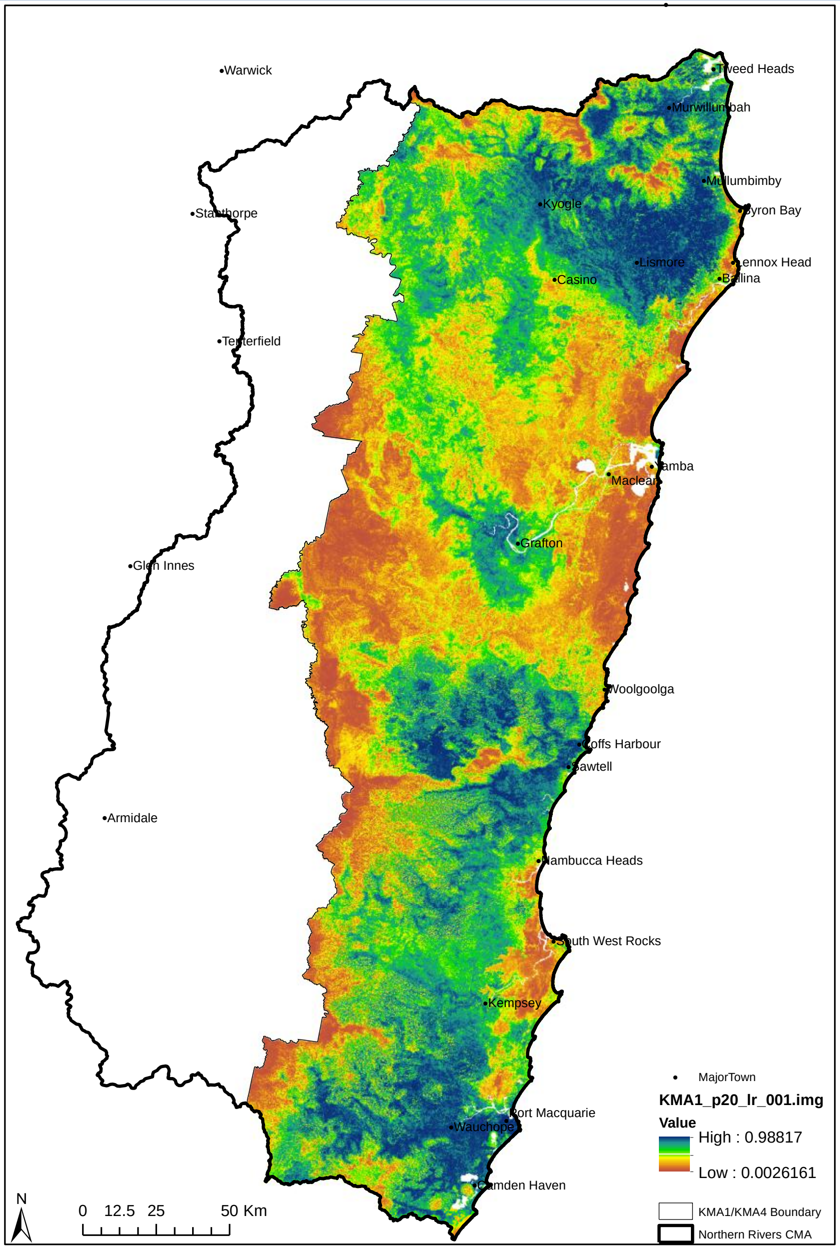
Example where two variables can be removed



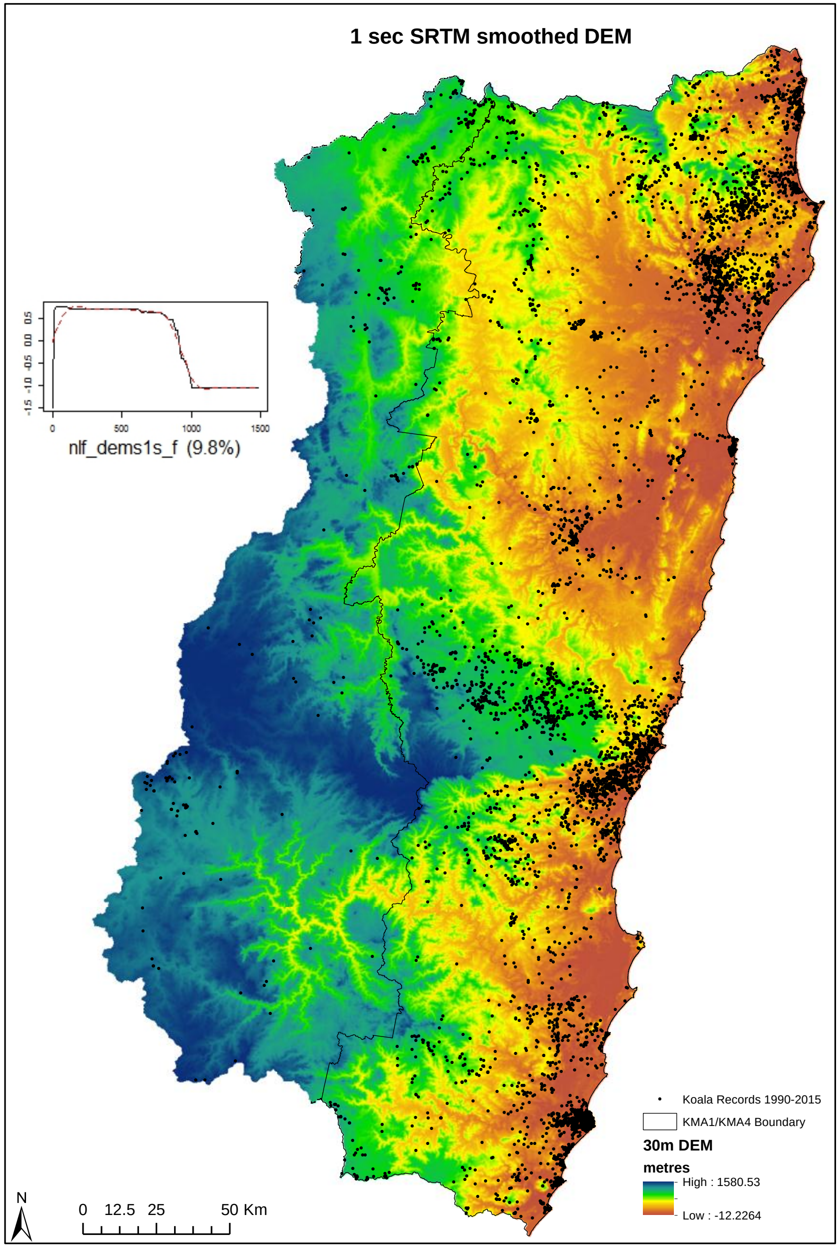
In this case, all 20 predictors retained

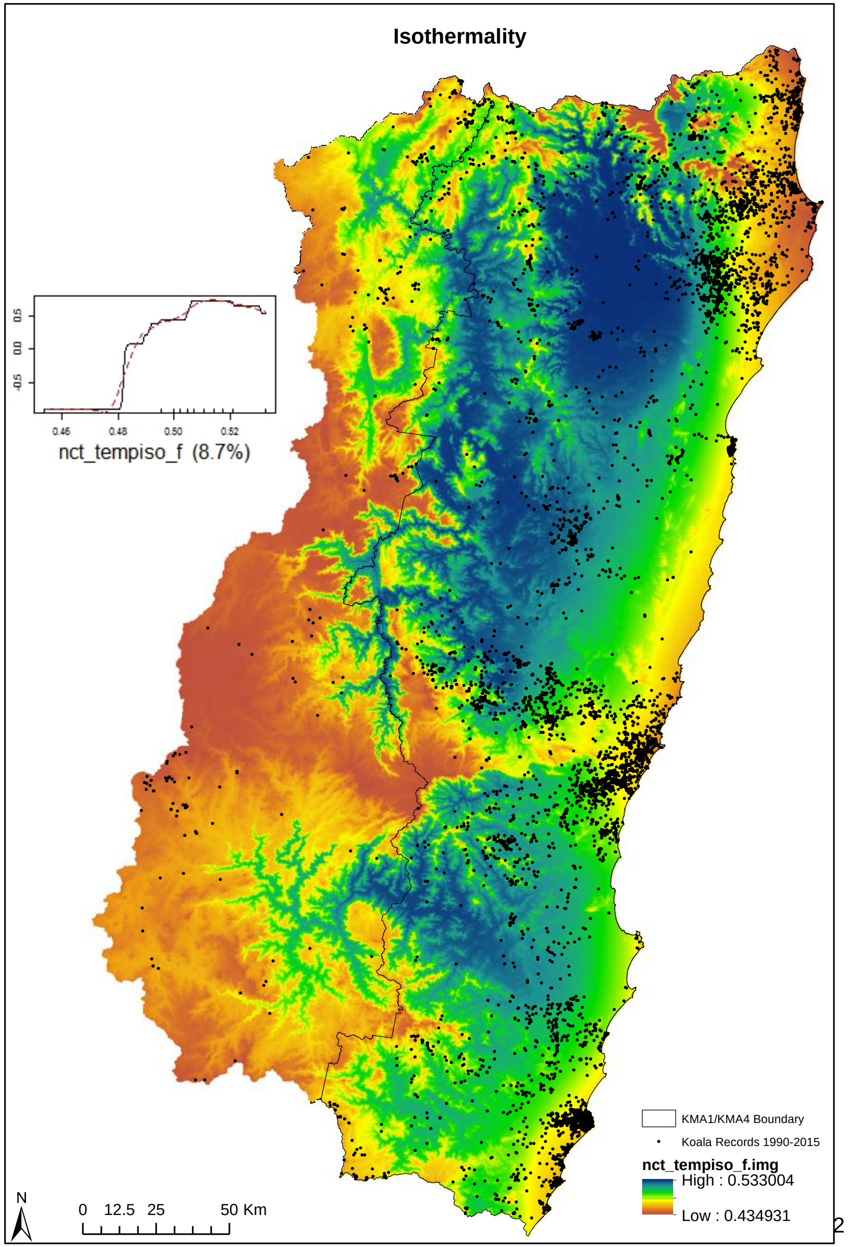


Probability of occurrence output from R

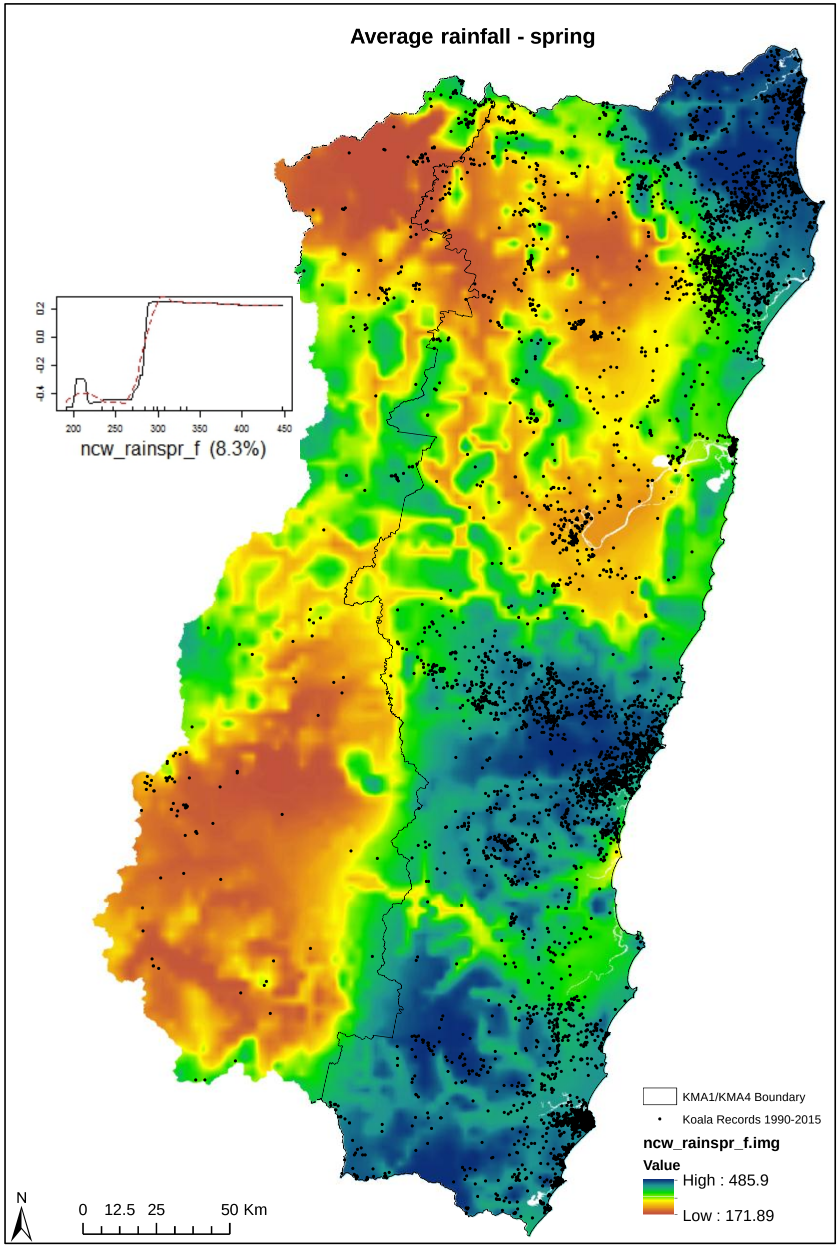


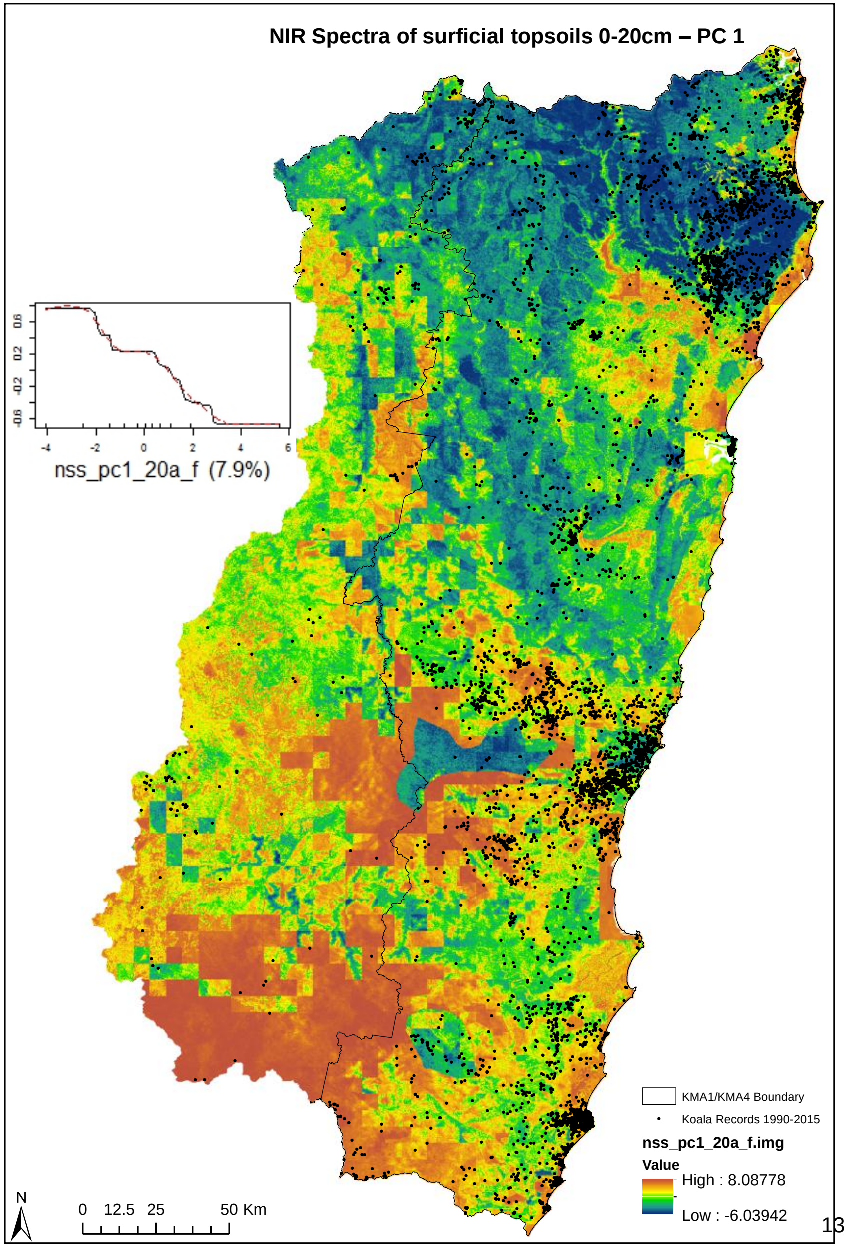
KMA1: Fitted functions for predictors



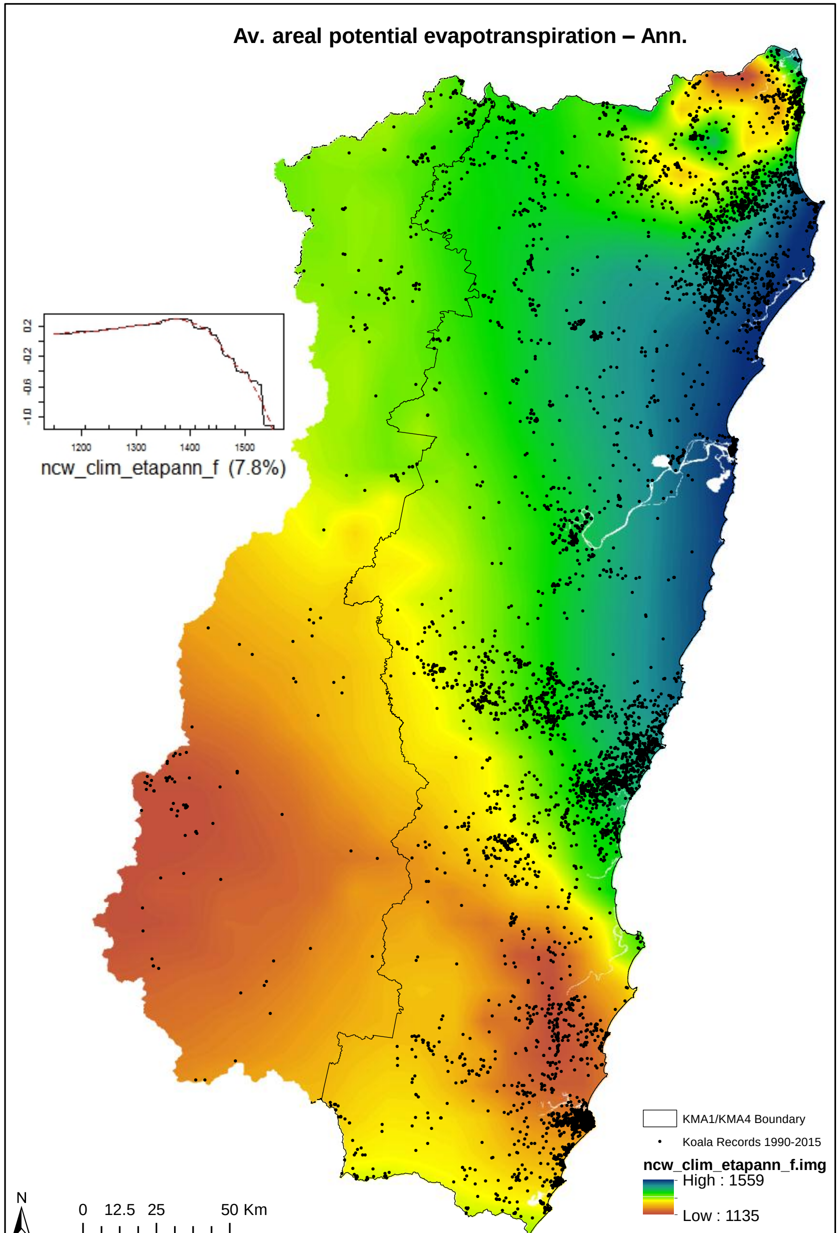


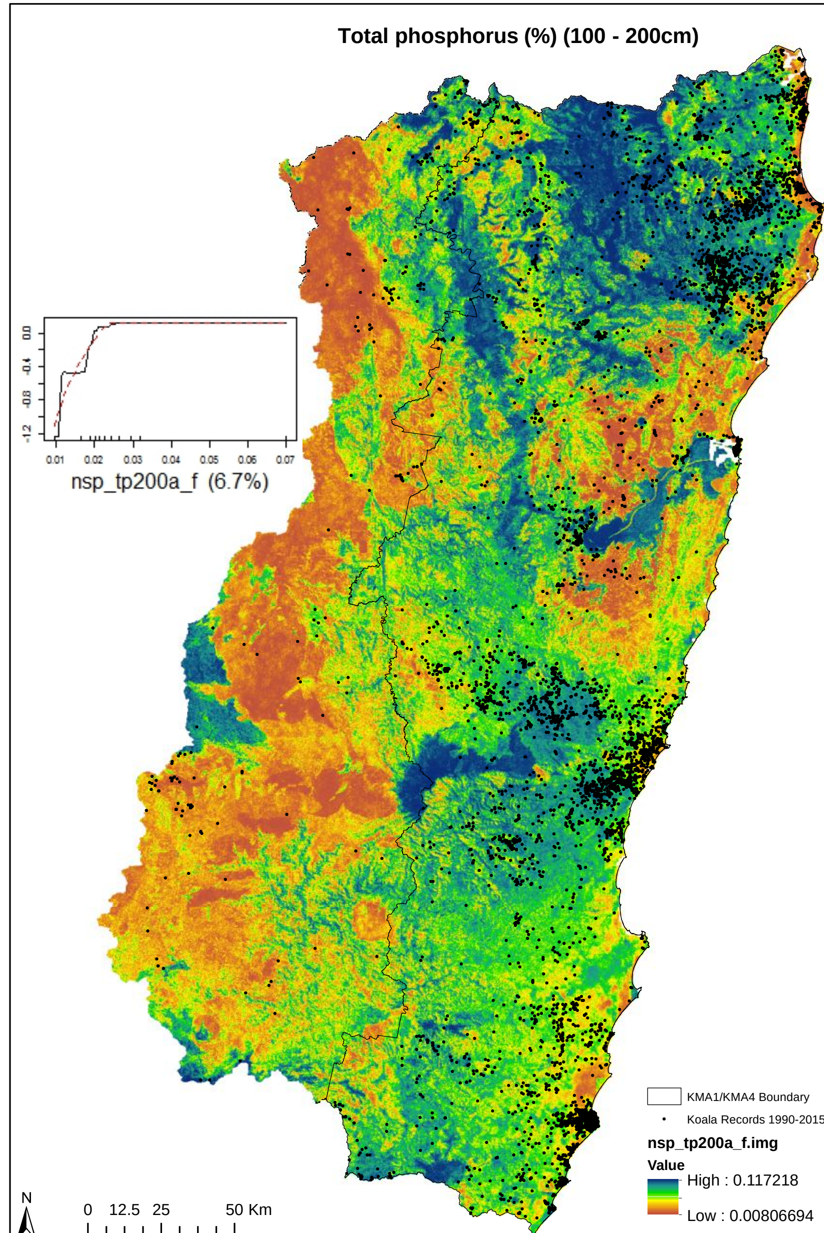
KMA1: Fitted functions for predictors



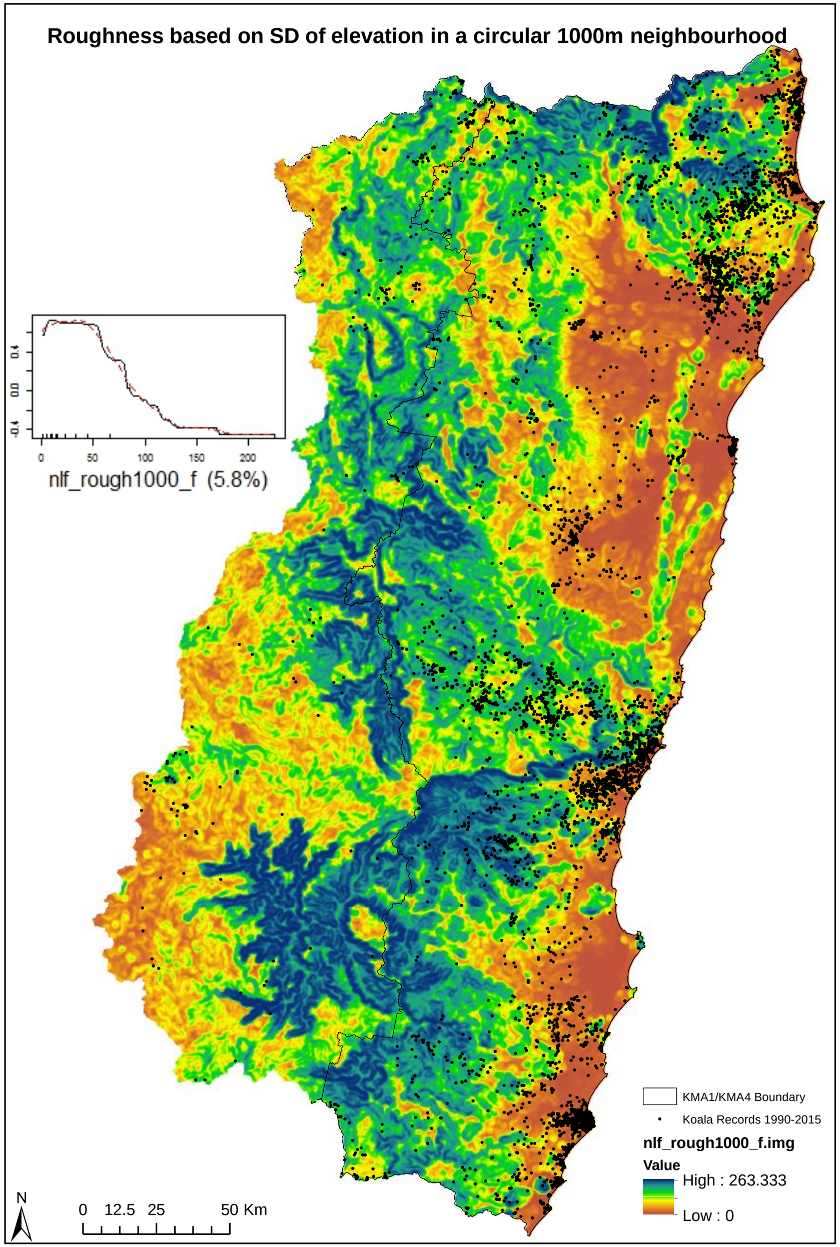


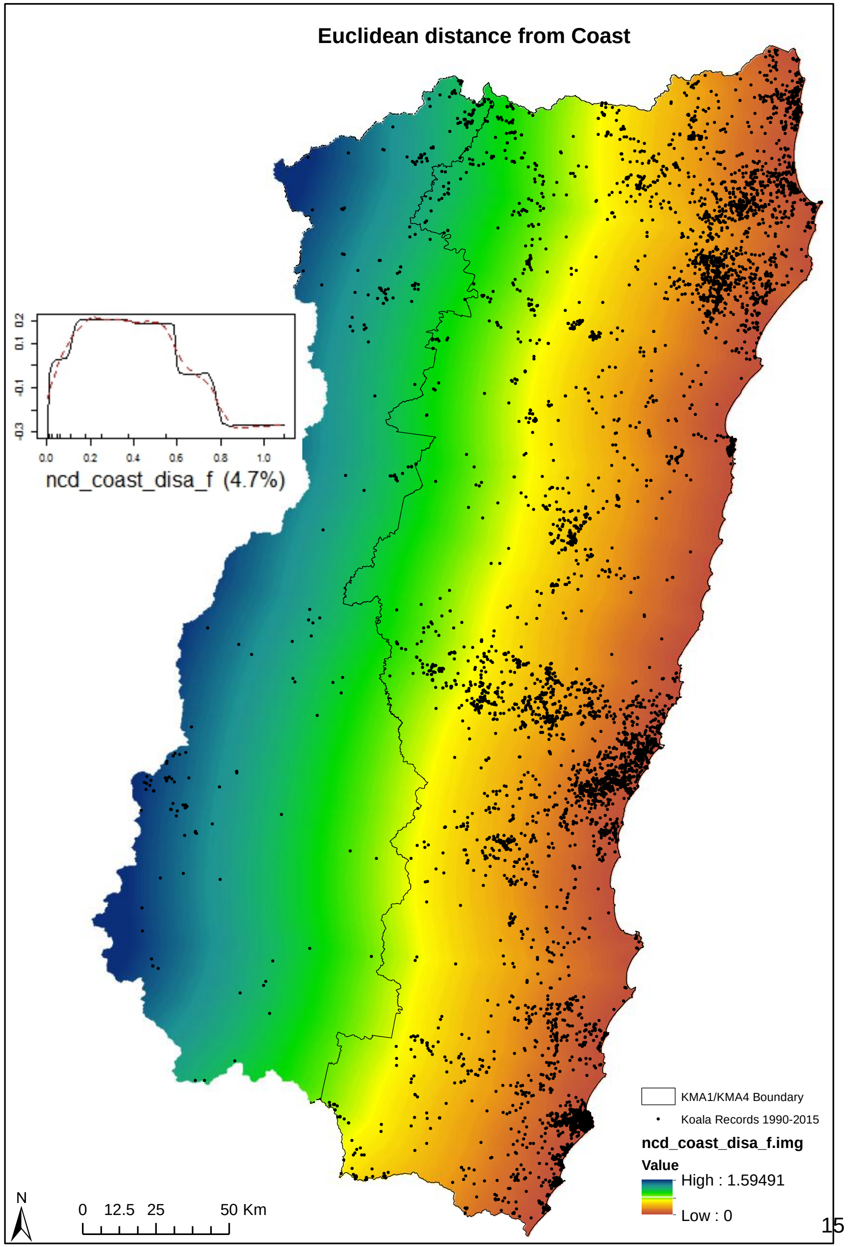
KMA1: Fitted functions for predictors



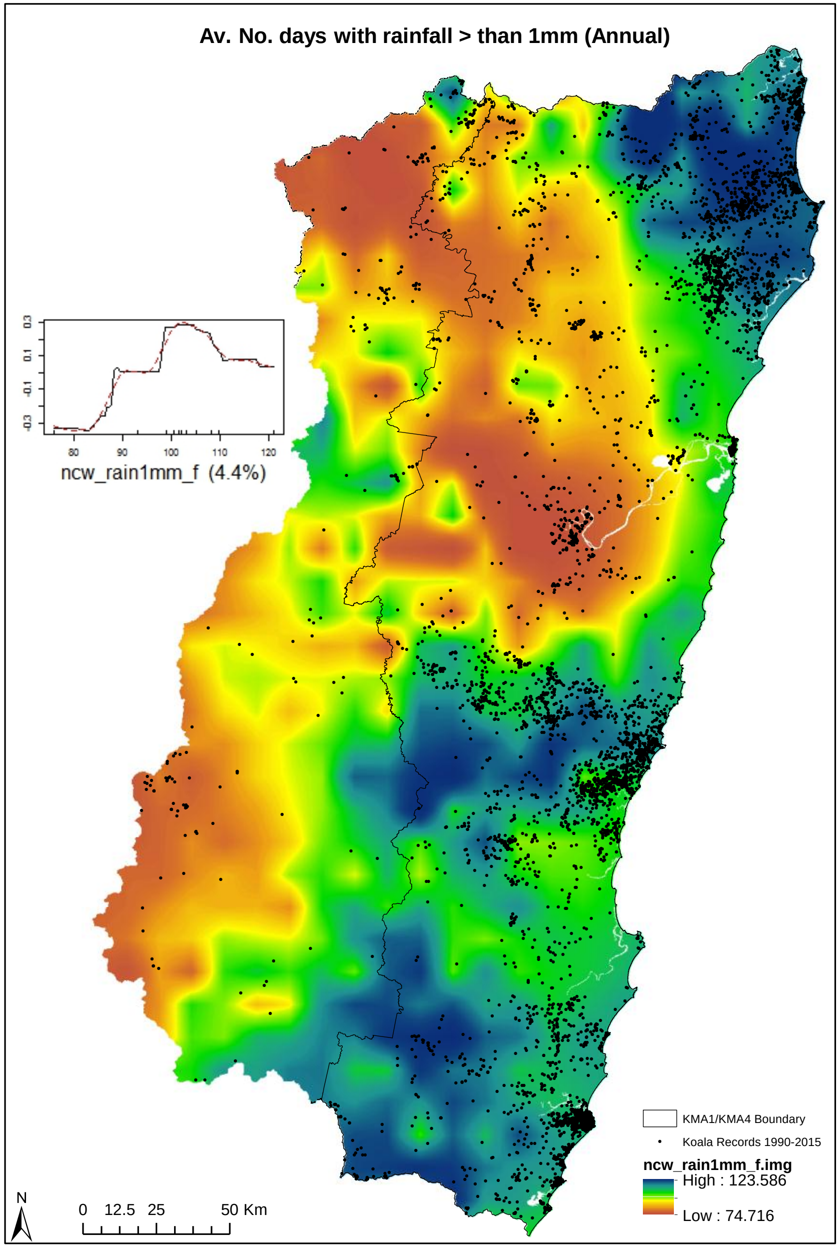


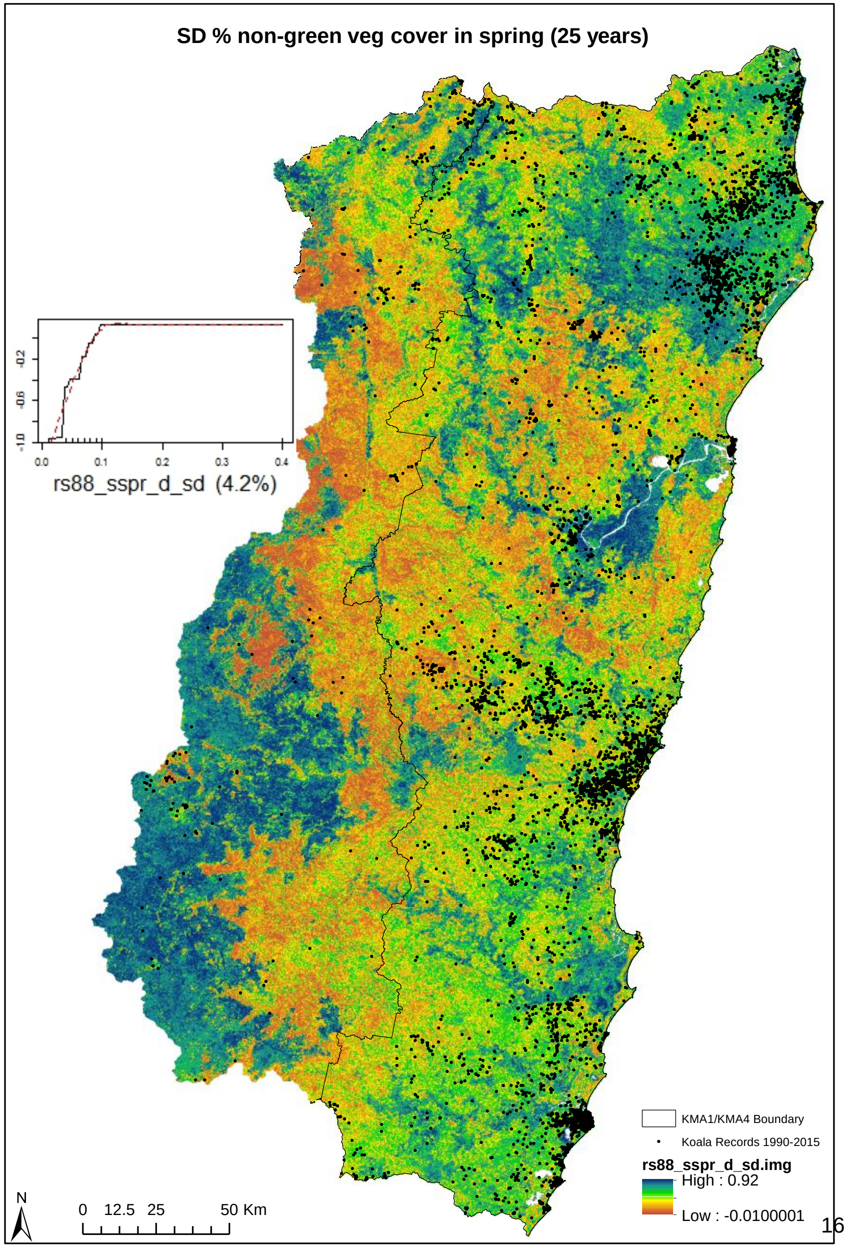
KMA1: Fitted functions for predictors



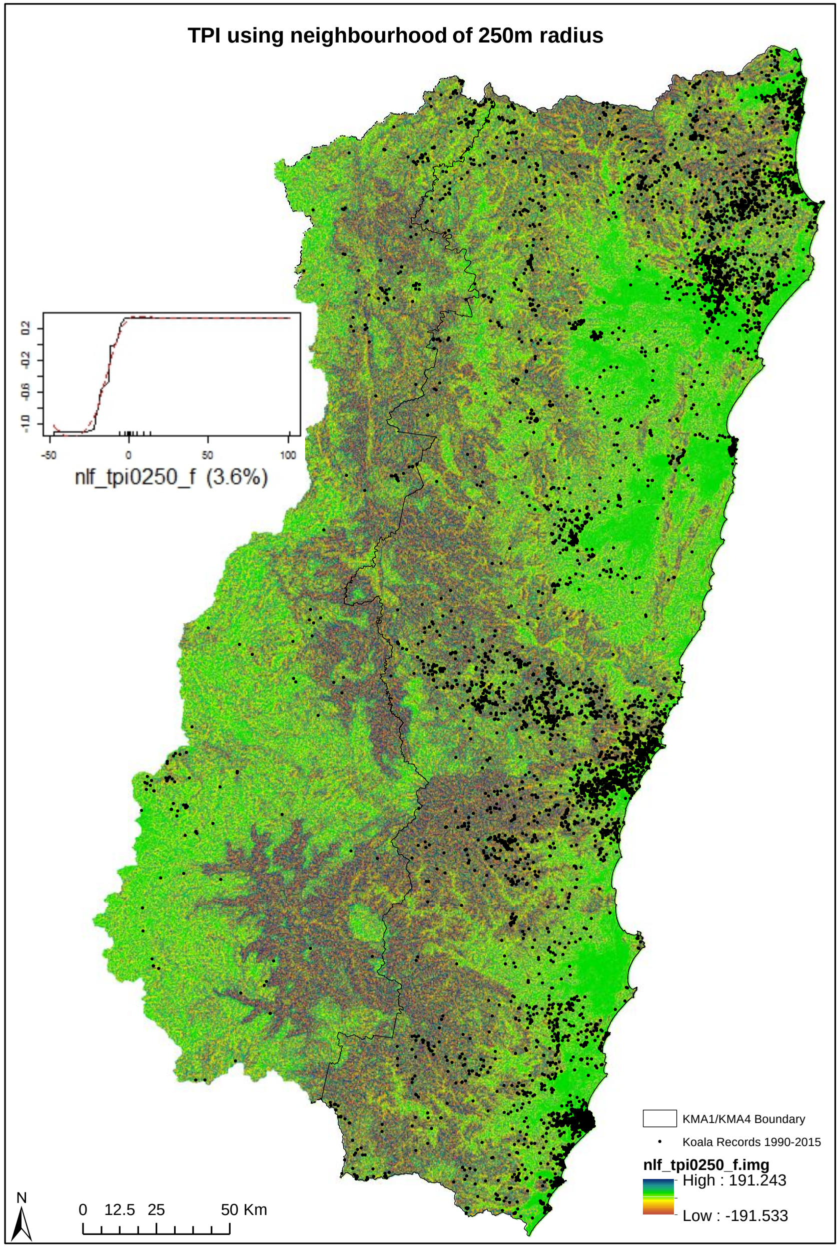


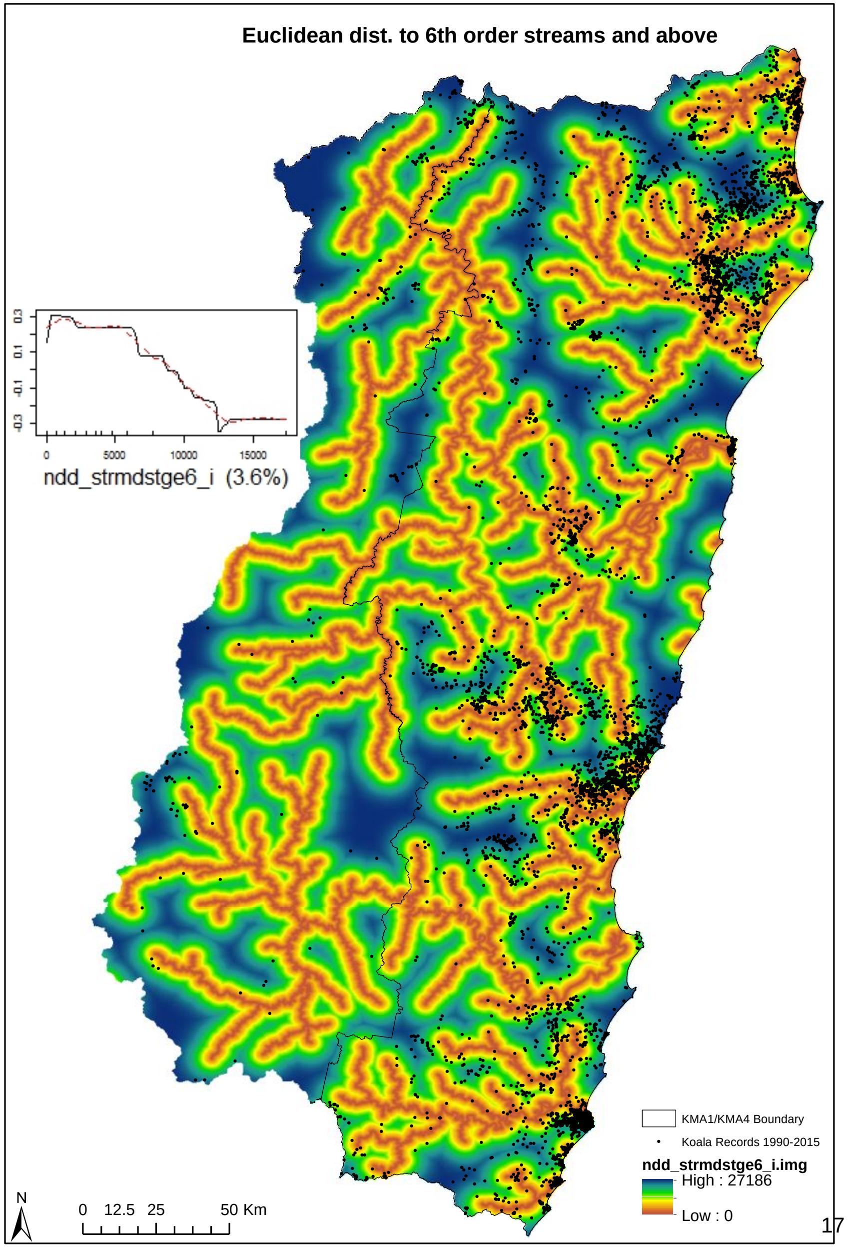
KMA1: Fitted functions for predictors



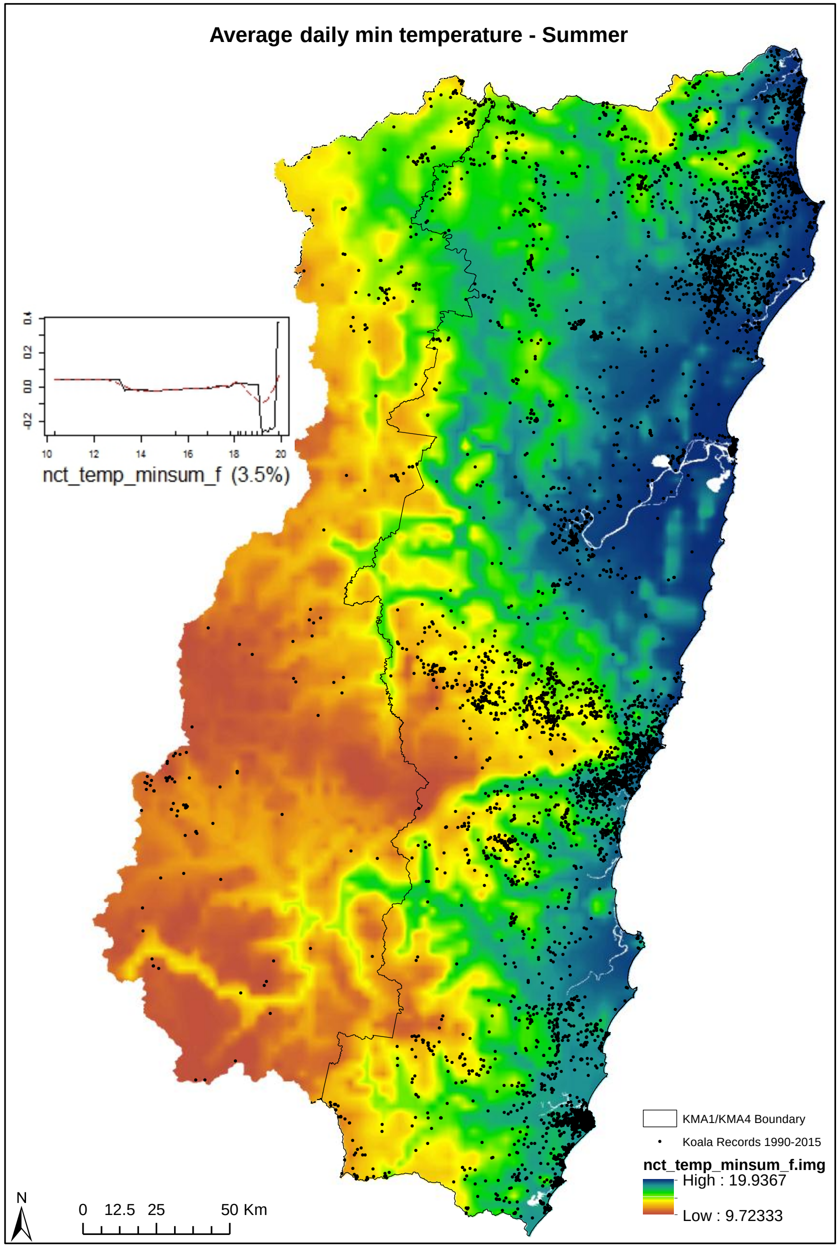


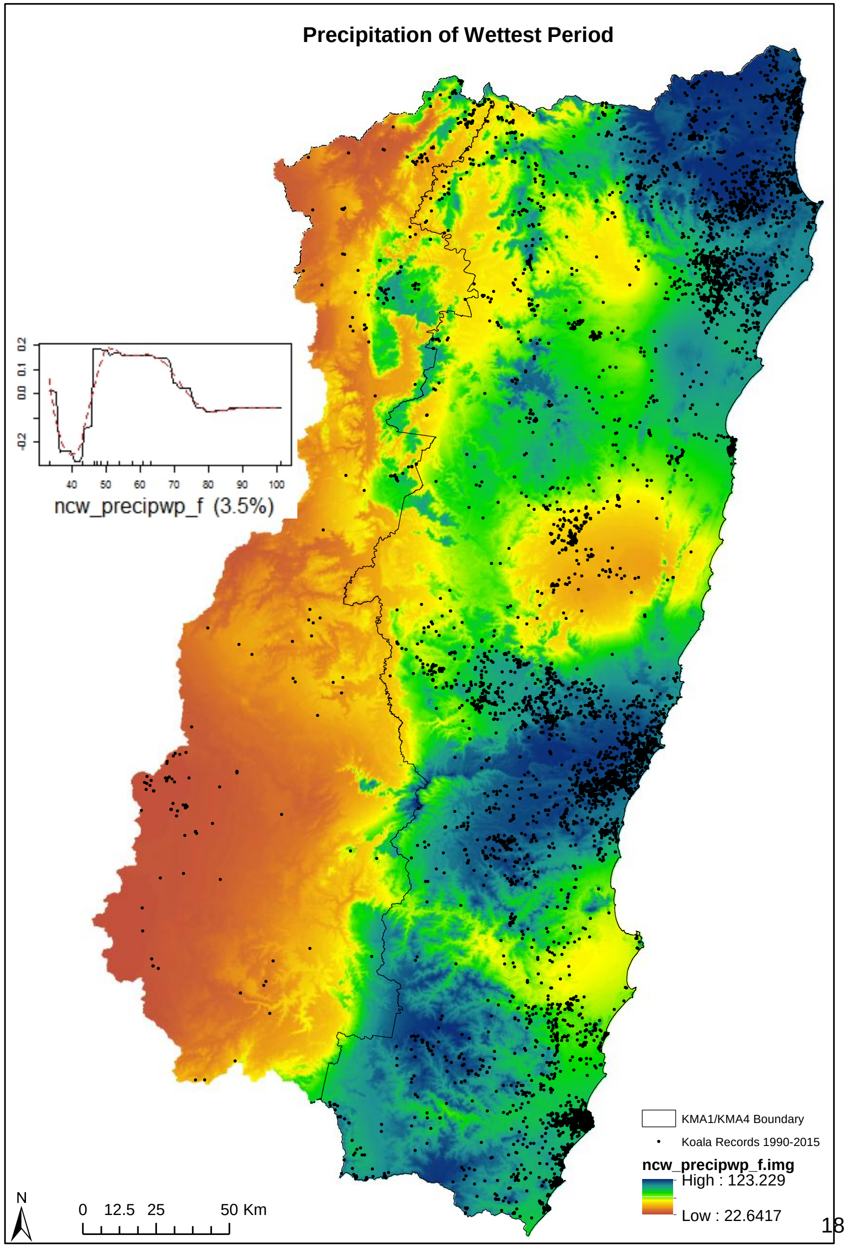
KMA1: Fitted functions for predictors



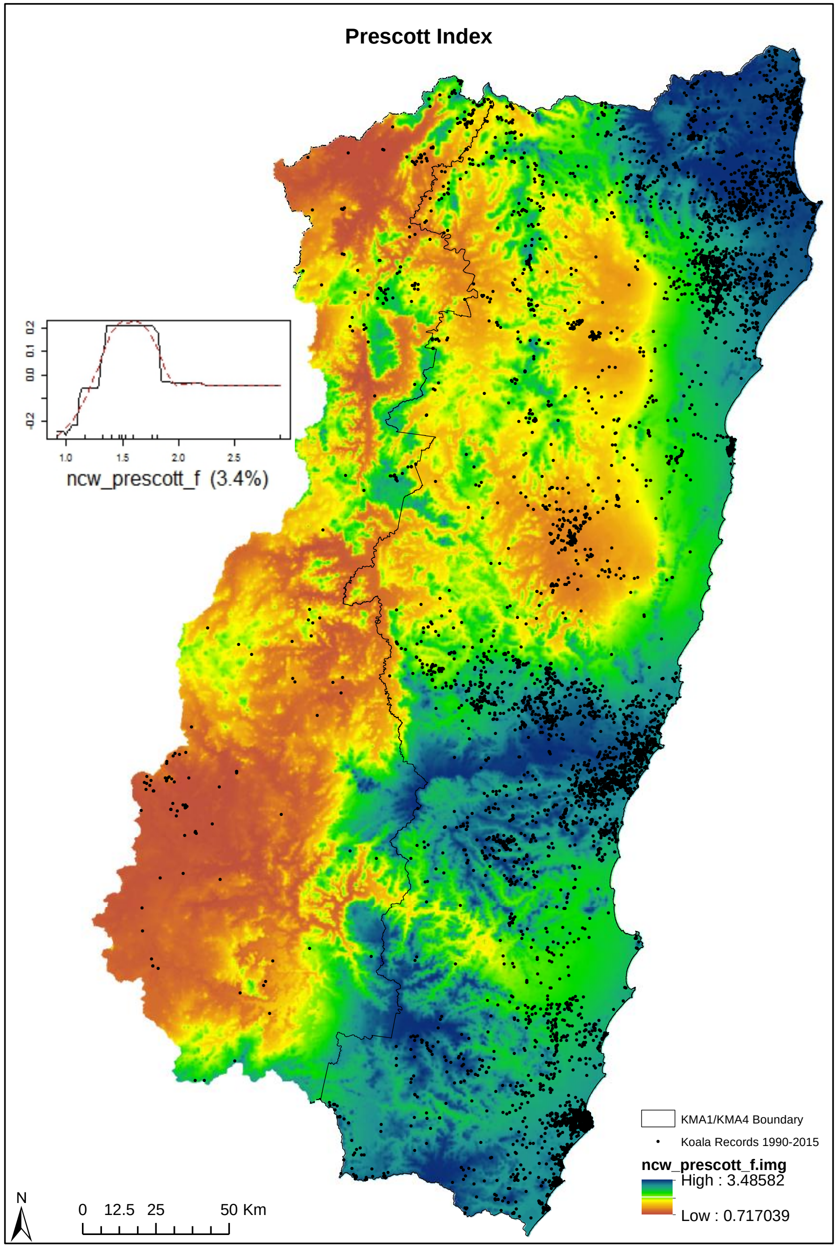


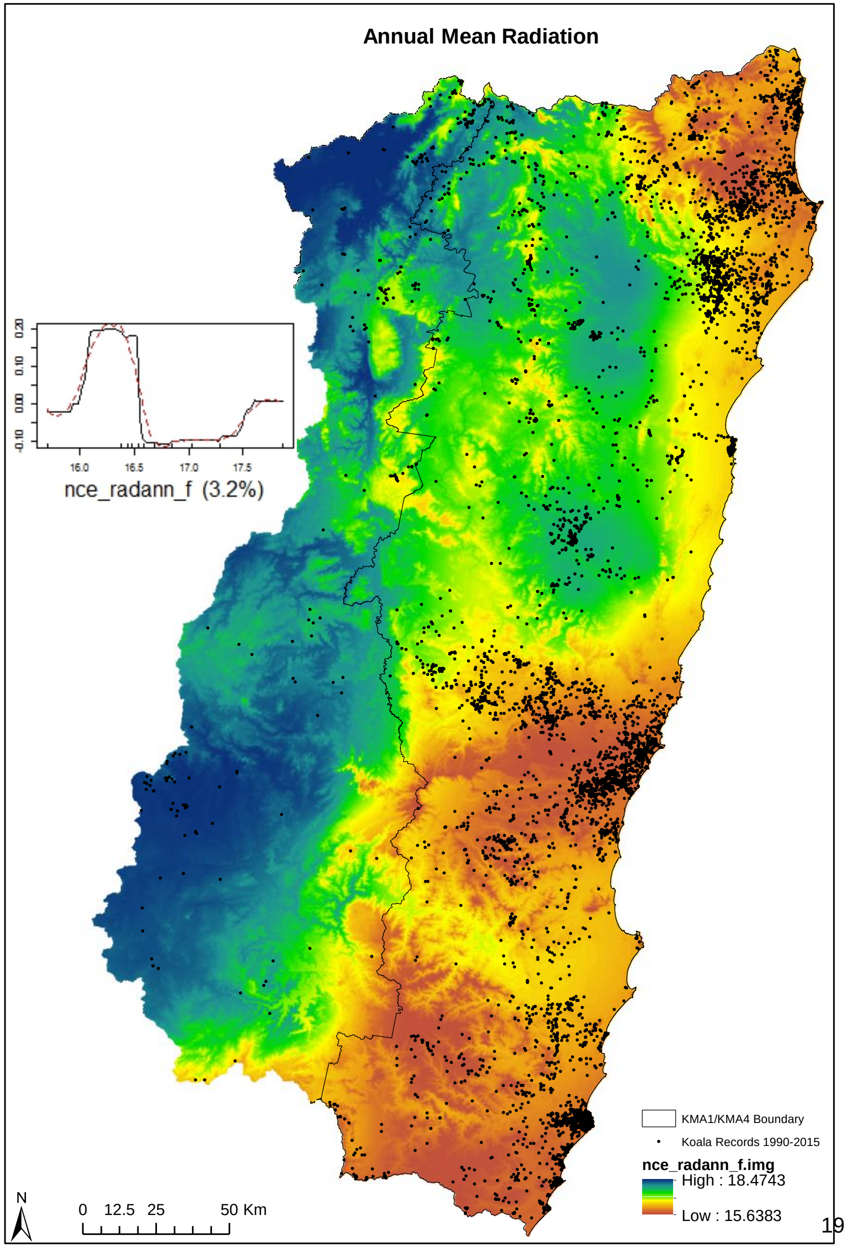
KMA1: Fitted functions for predictors



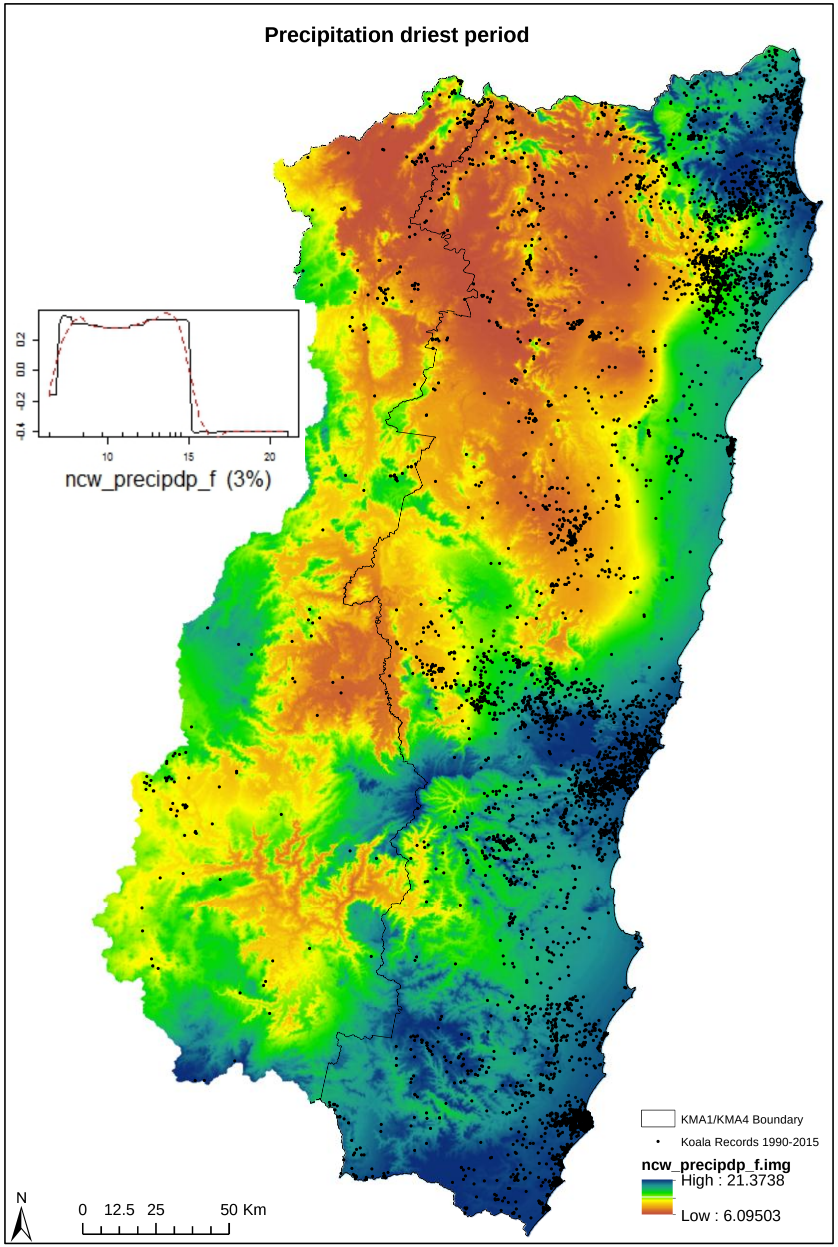


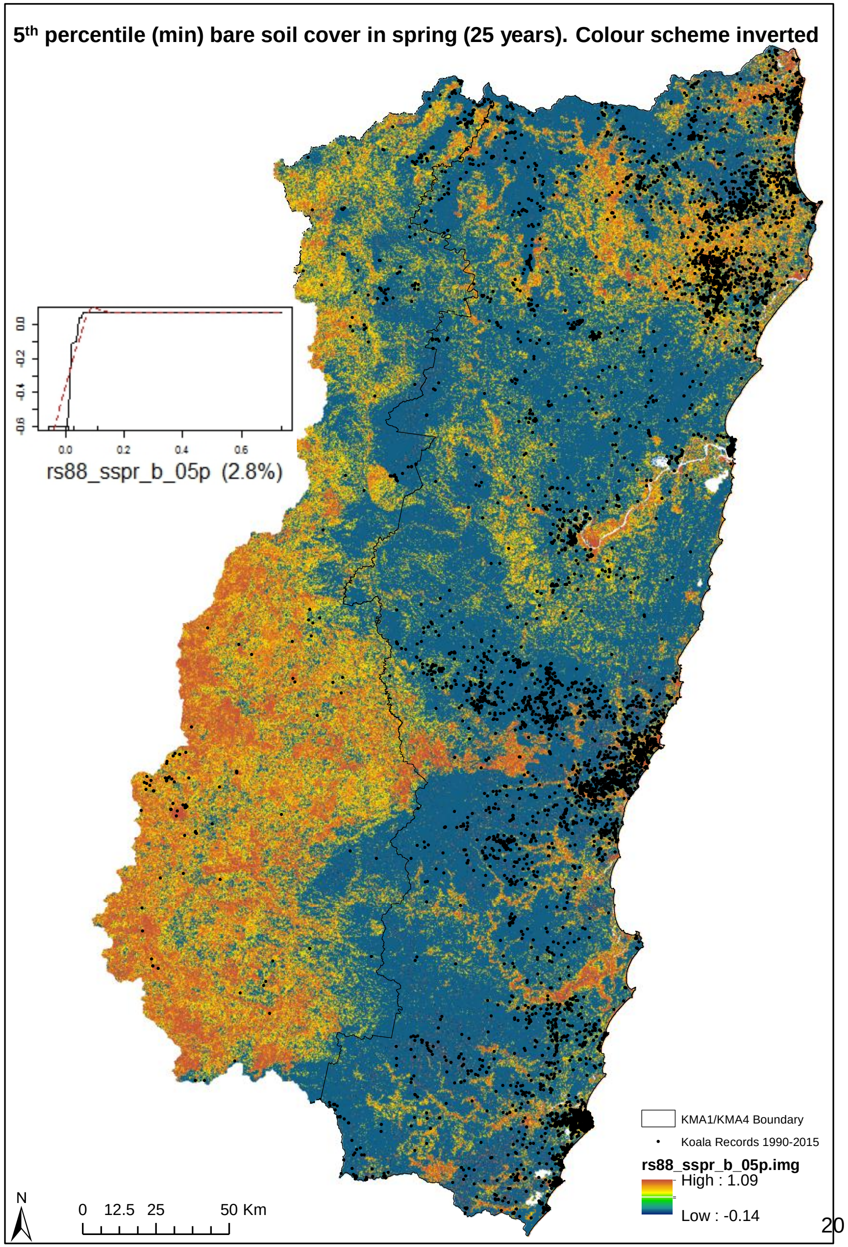
KMA1: Fitted functions for predictors





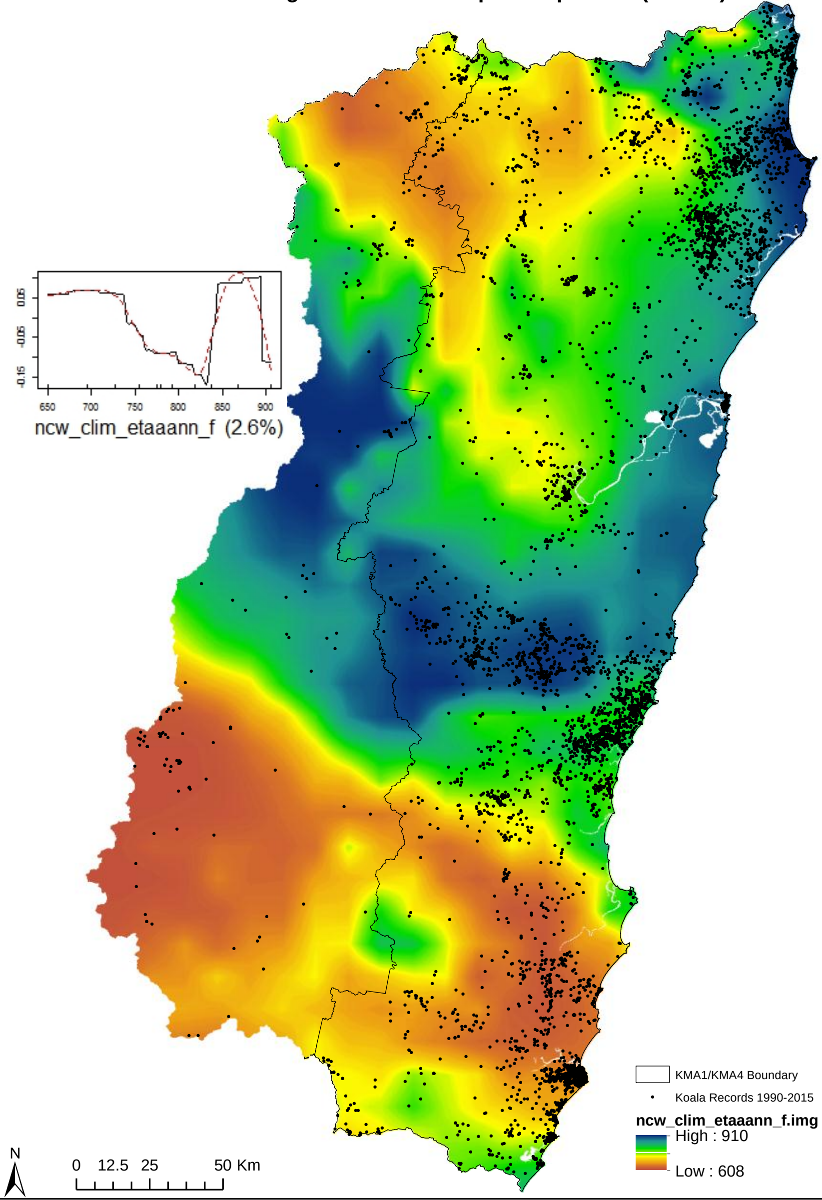
KMA1: Fitted functions for predictors



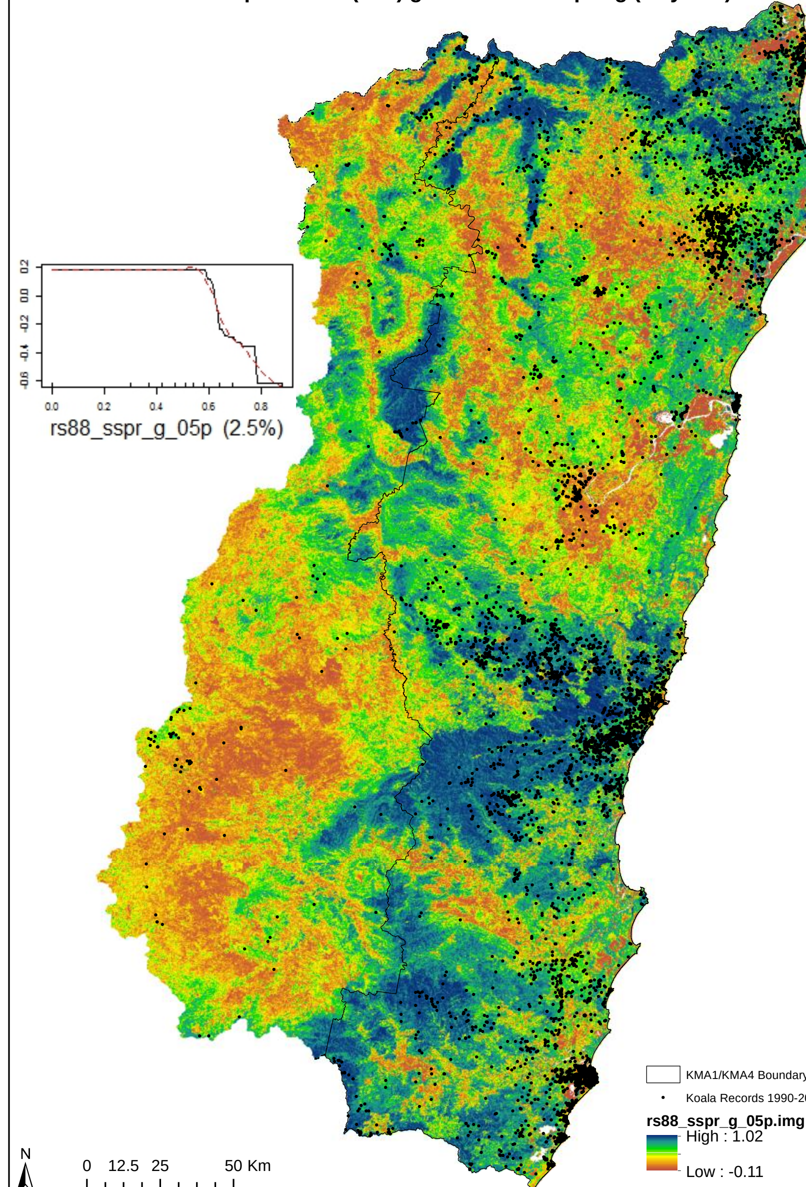


KMA1: Fitted functions for predictors

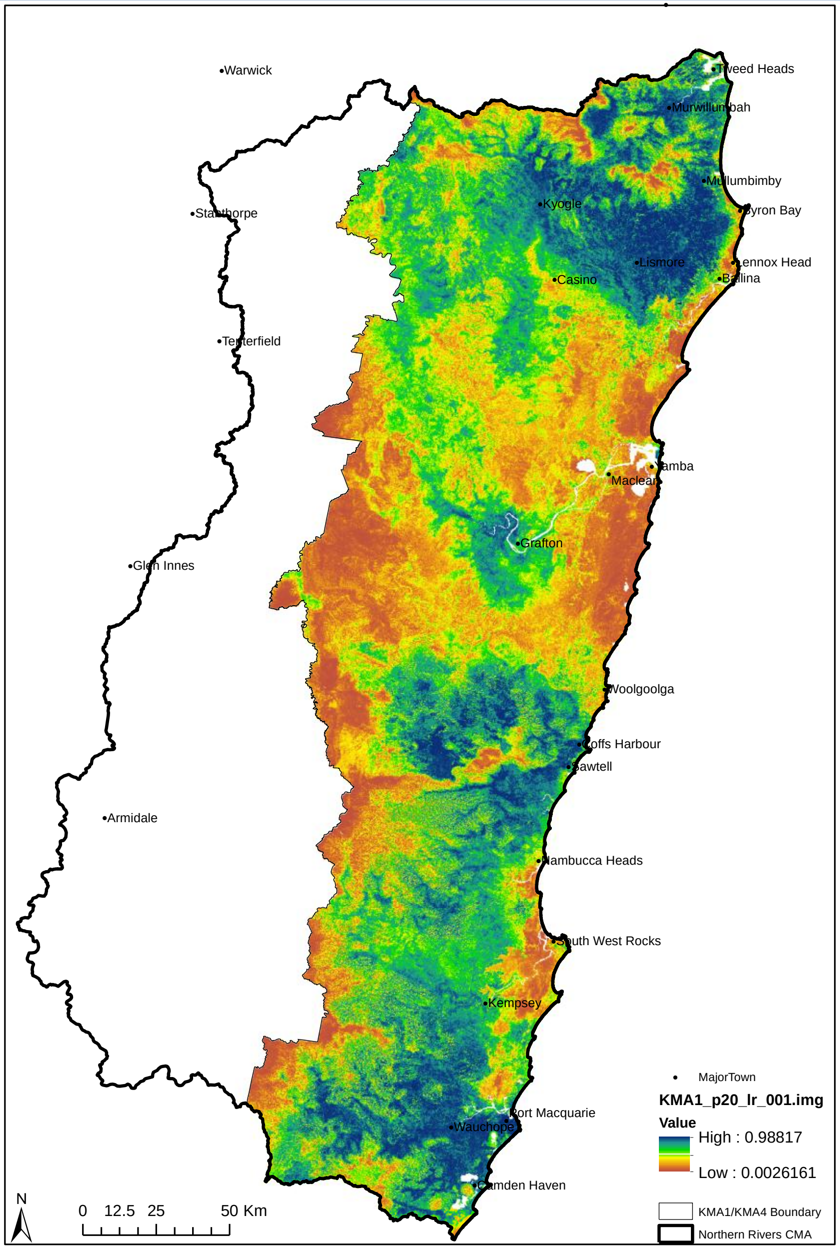
Average areal actual evapotranspiration (Annual).



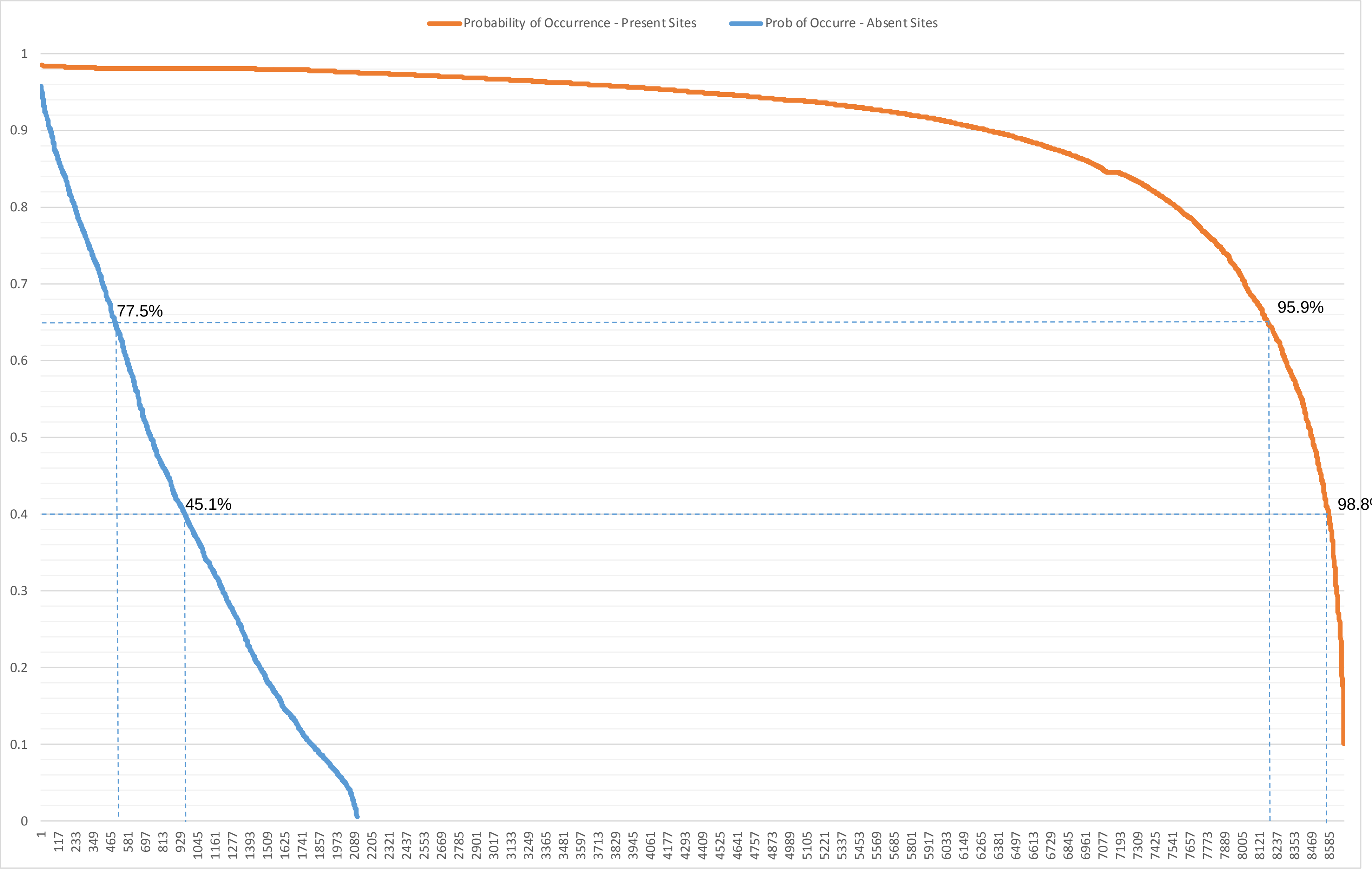
5th percentile (min) green cover in spring (25 years)



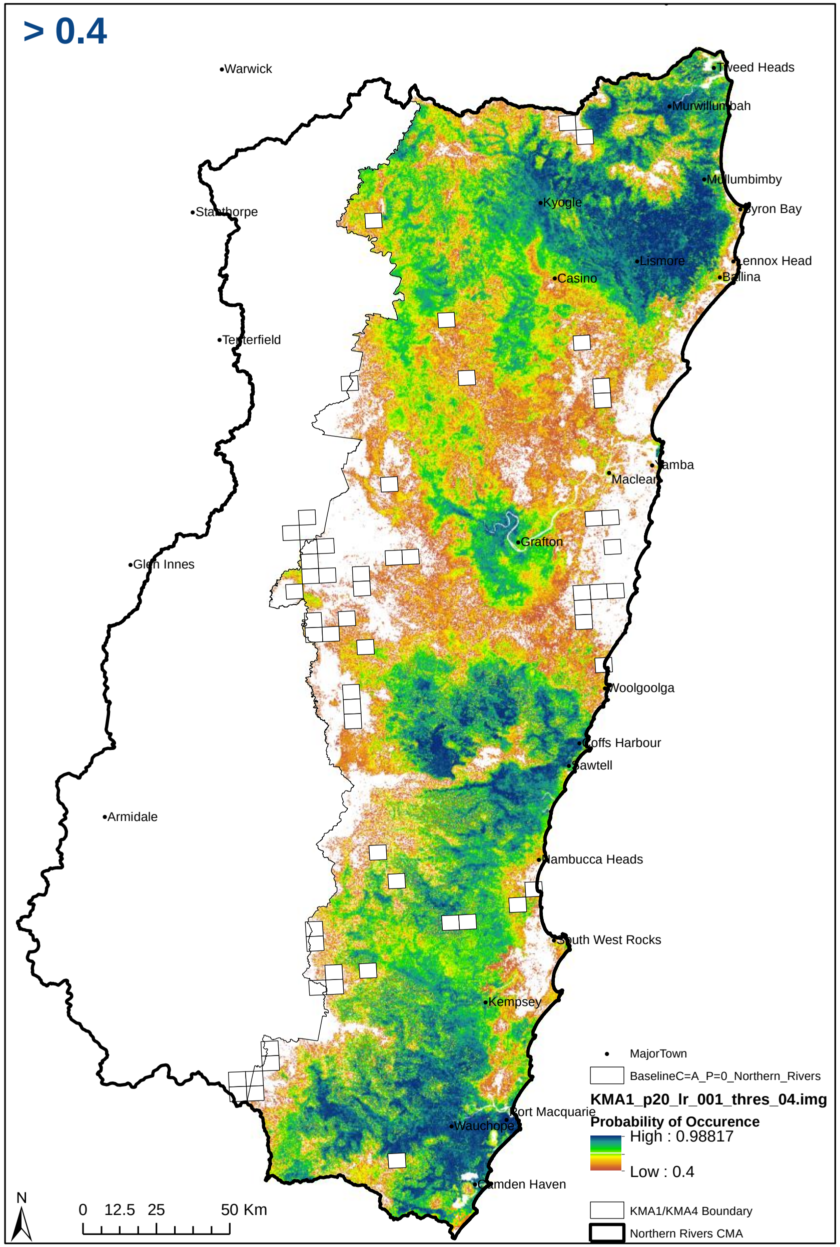
Criteria for defining suitable habitat

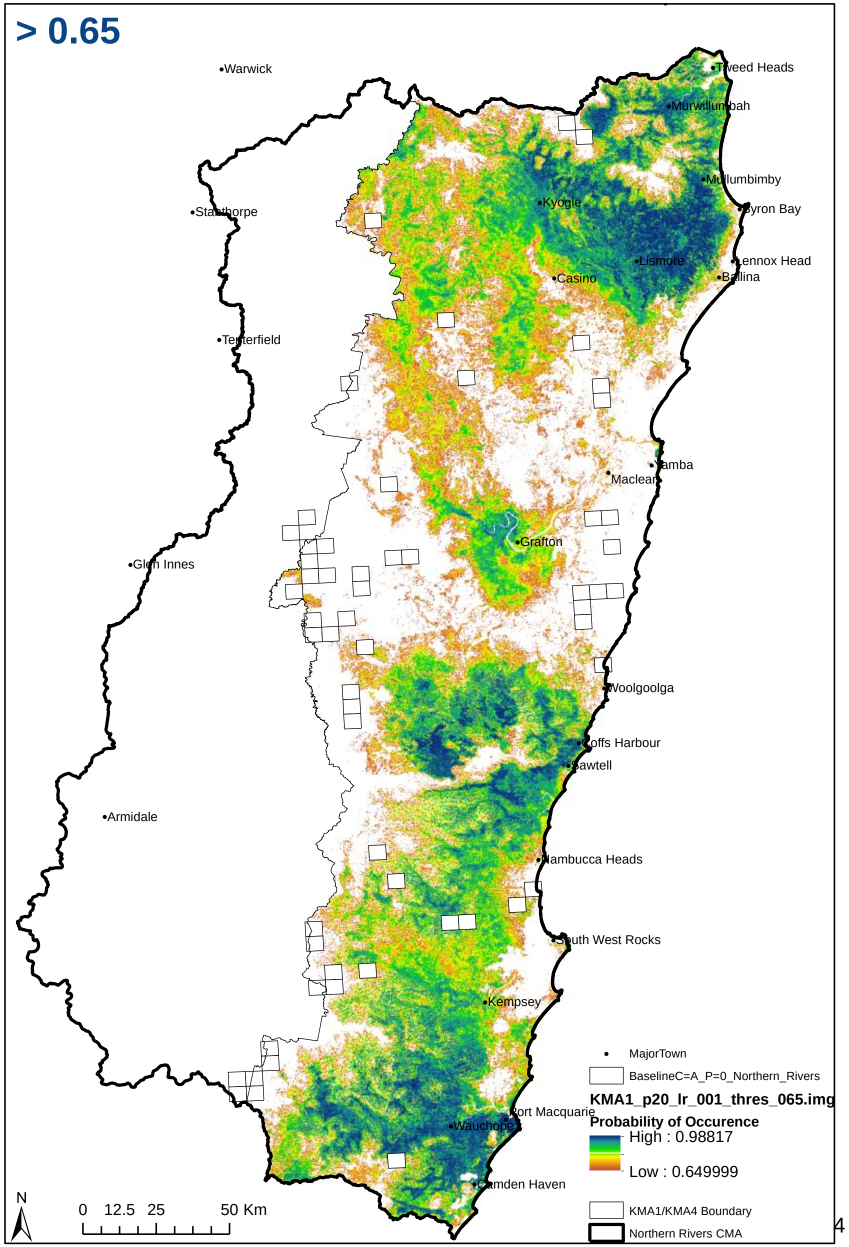


Performance Evaluation: Fitted Values



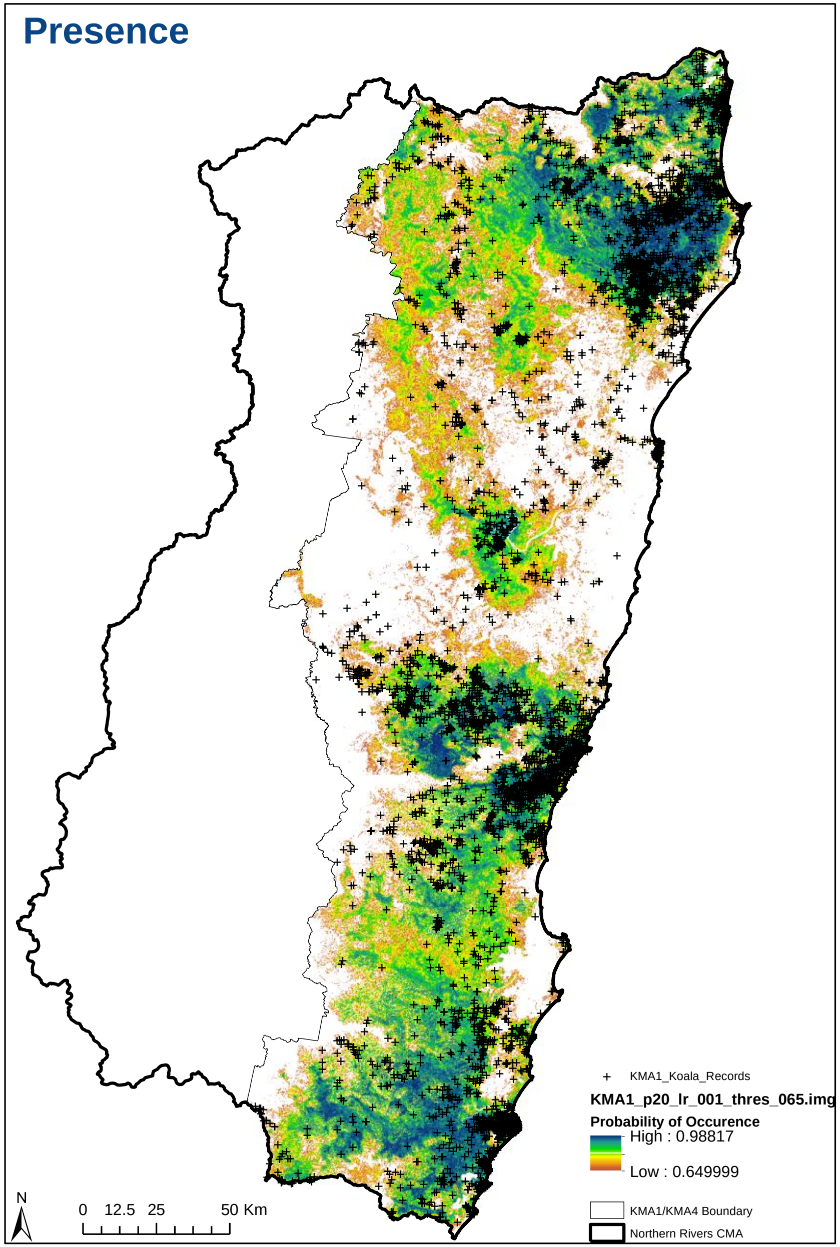
Habitat maps based on thresholds of 0.4 and 0.65



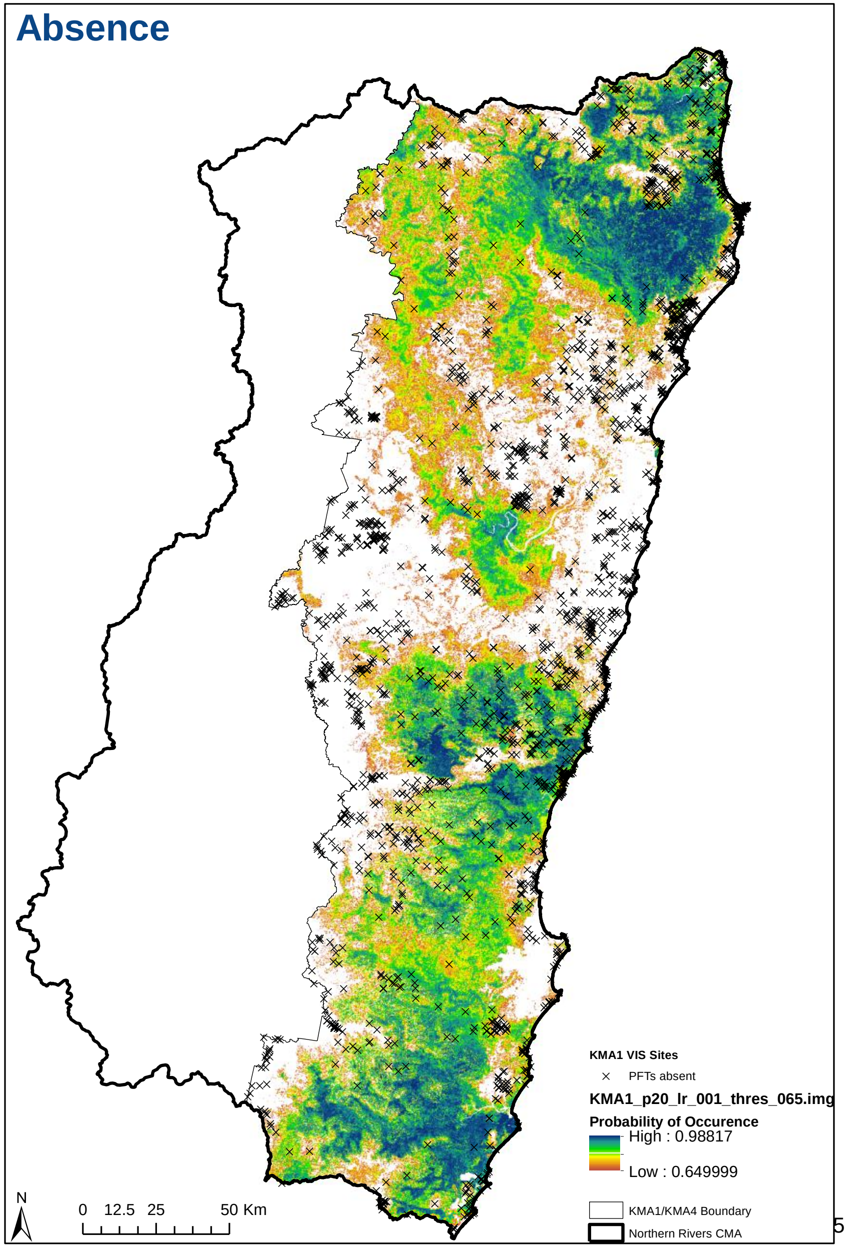


Habitat map (>0.65 threshold) overlaid with presence and absence records

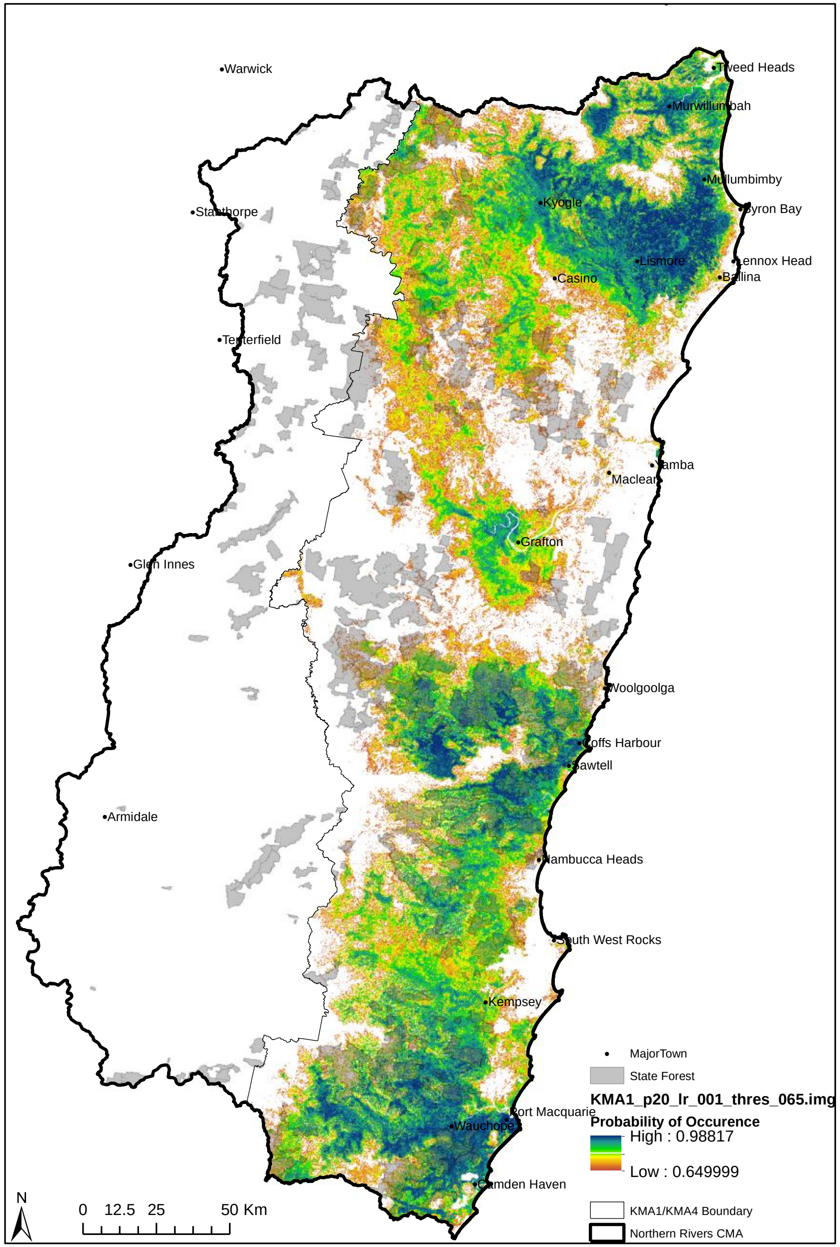
Presence

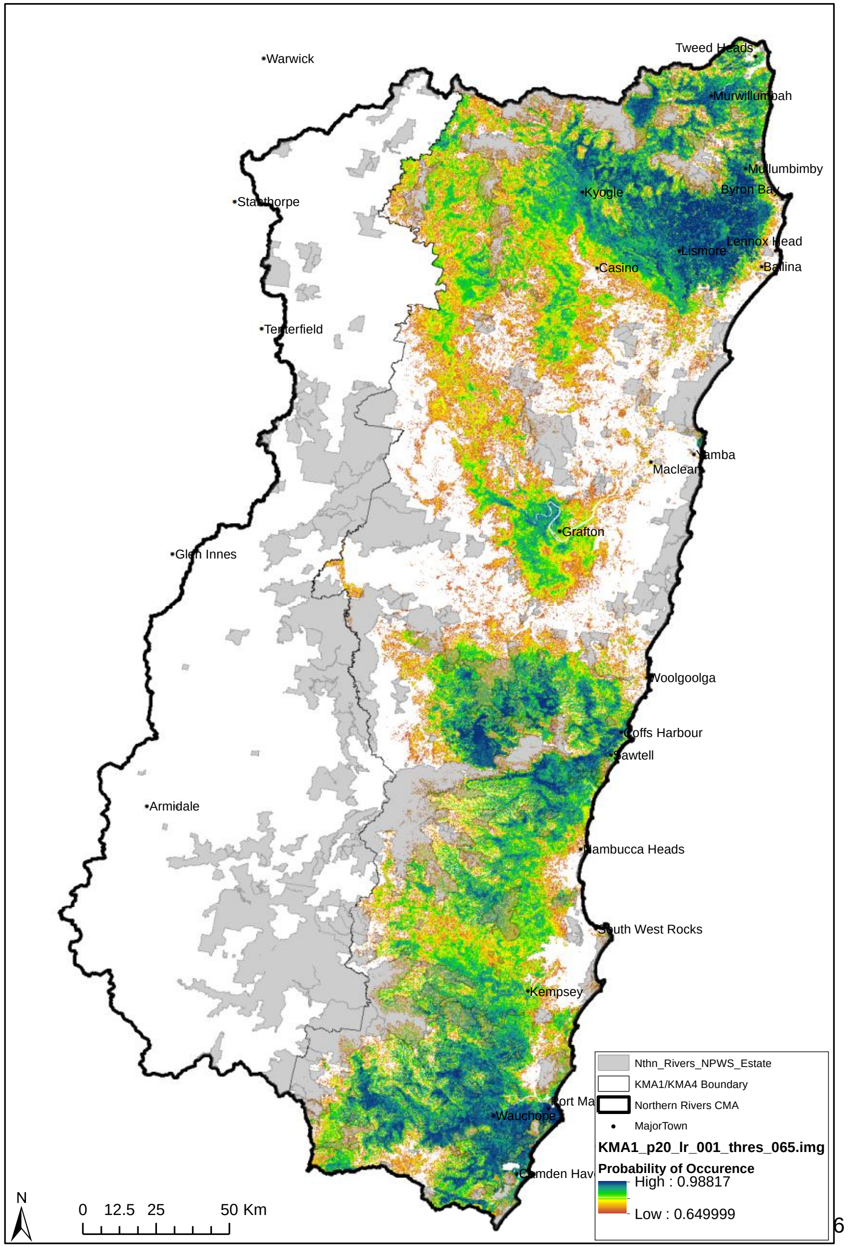


Absence

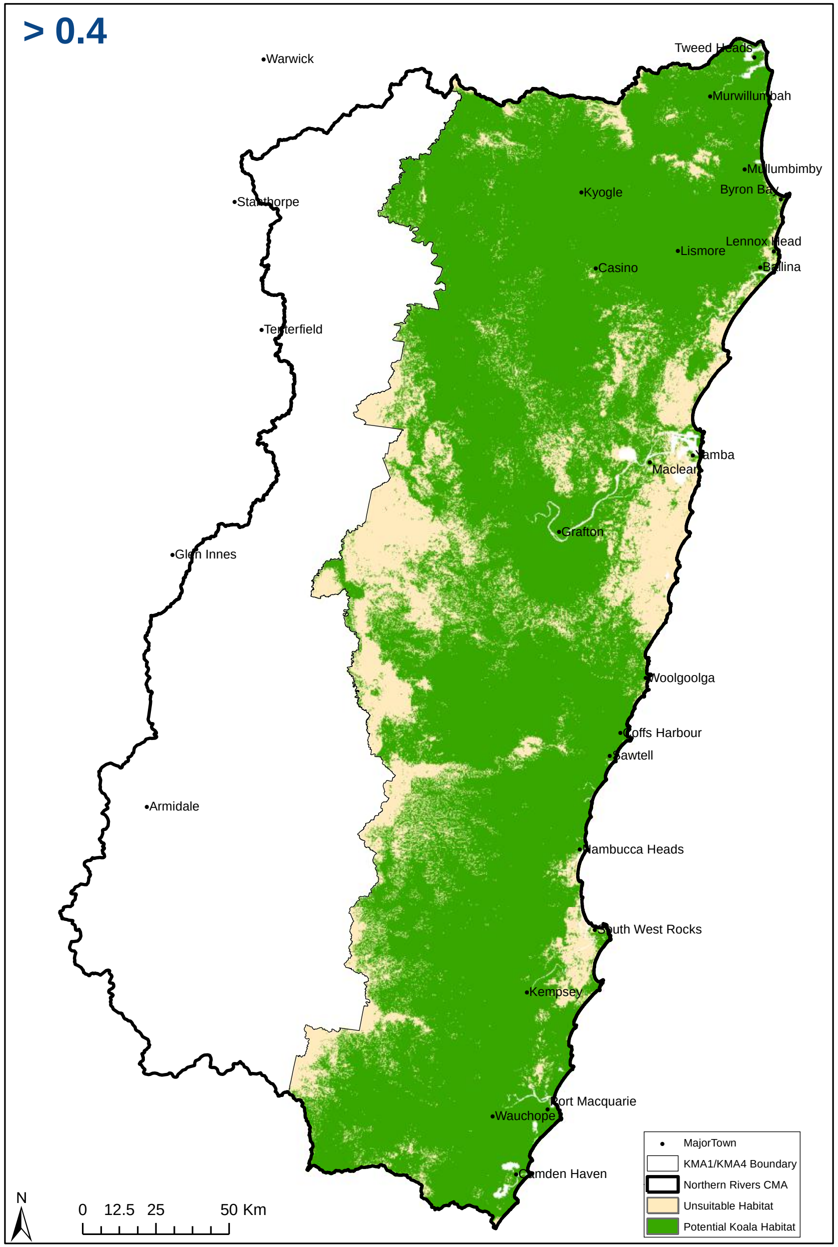


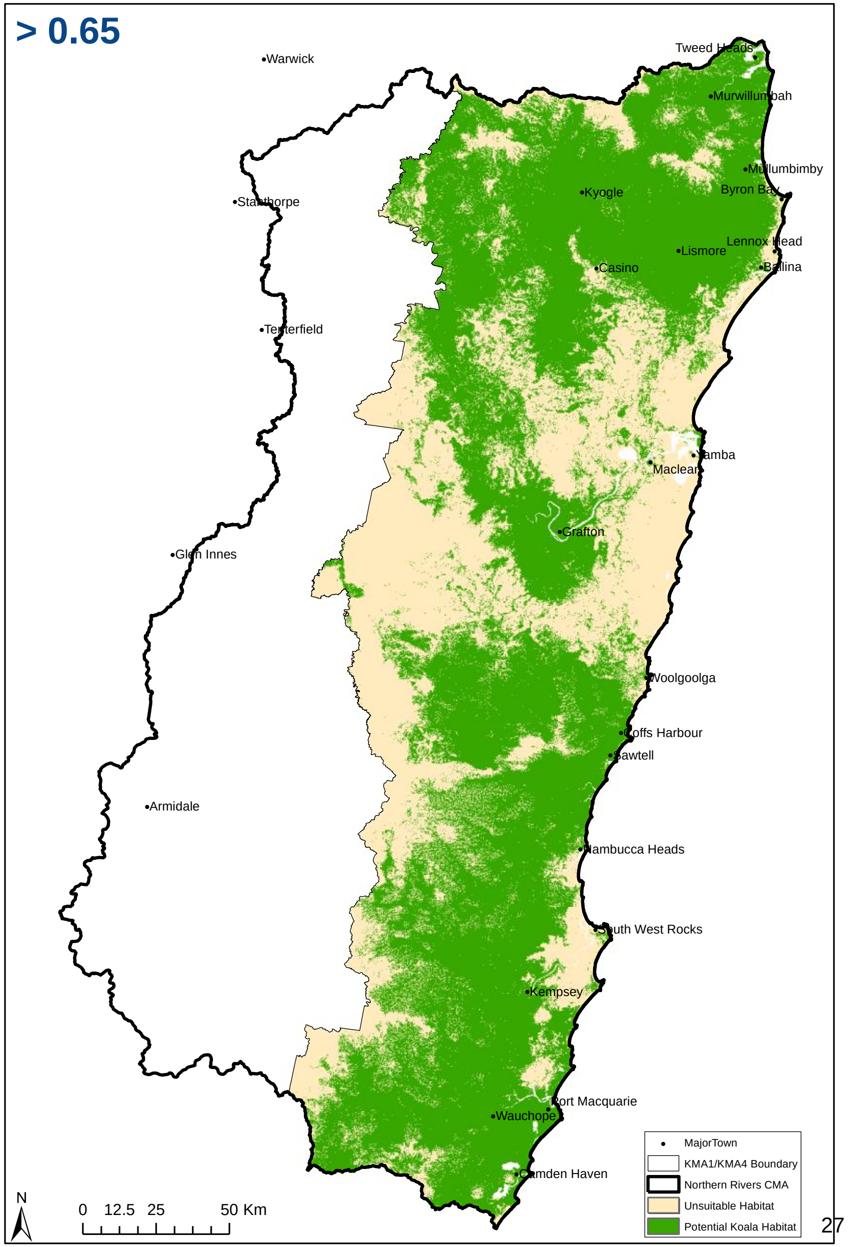
Habitat map in relation to State Forests and National Parks



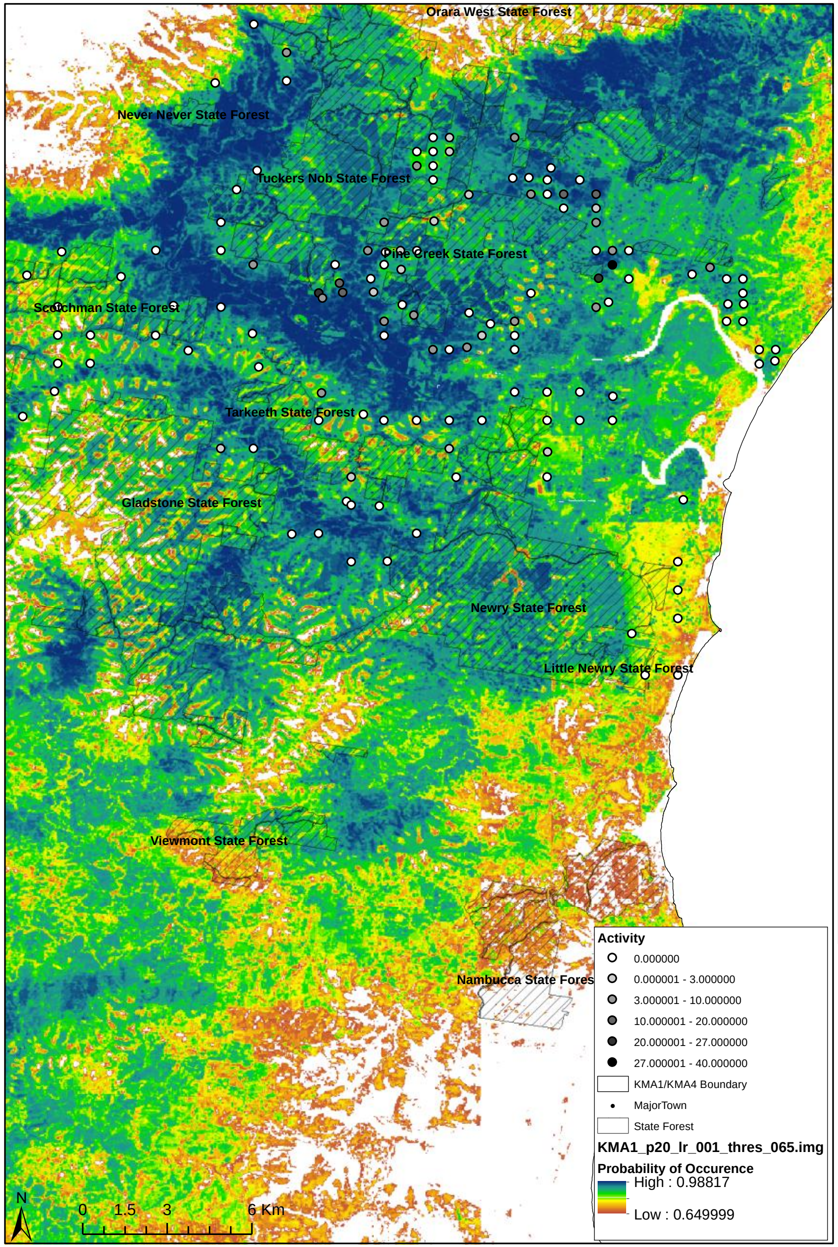


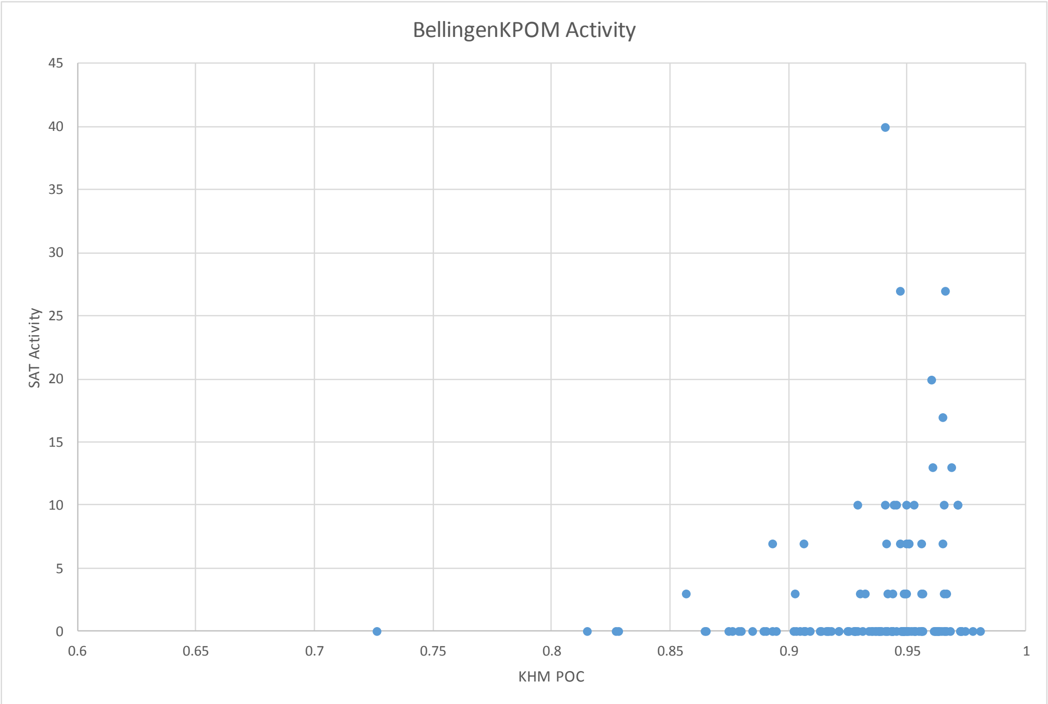
Unsuitable vs Potential Koala Habitat



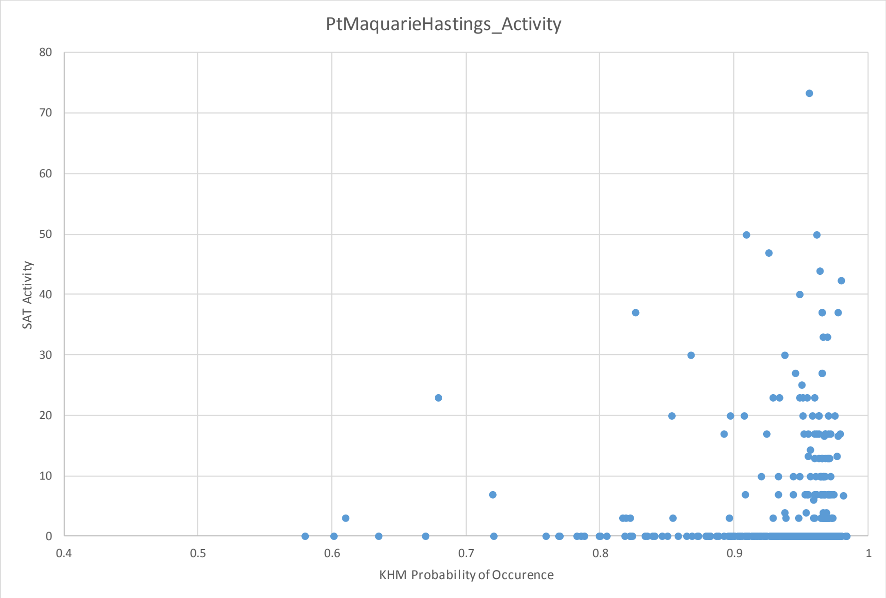


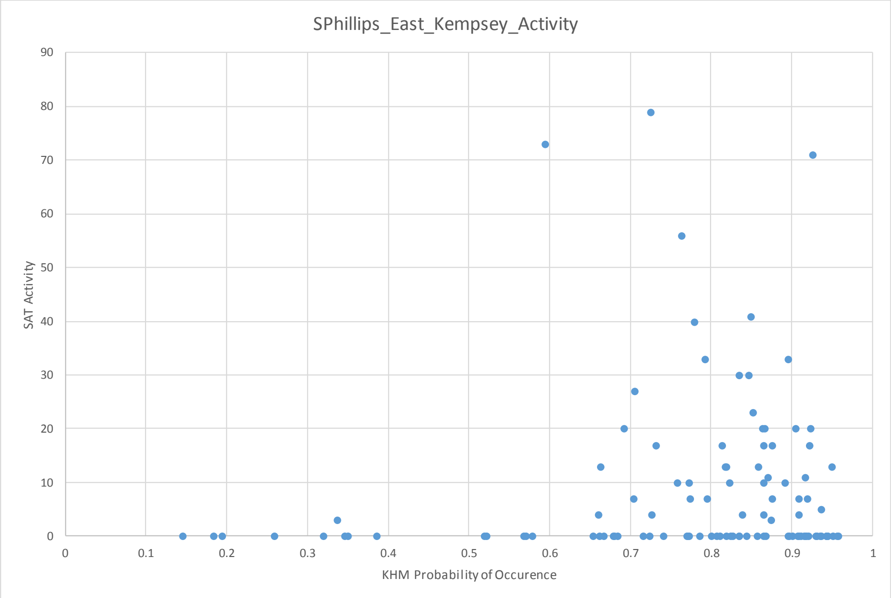
Predicted occurrence vs SAT Activity

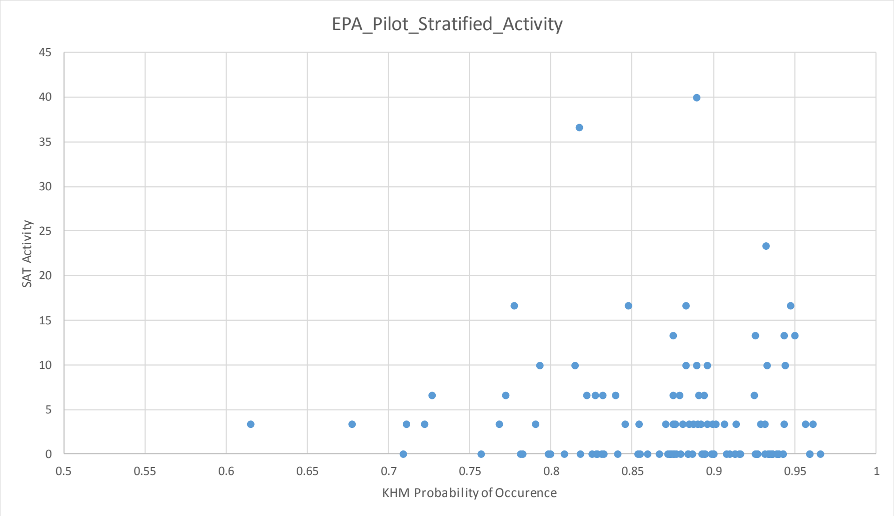


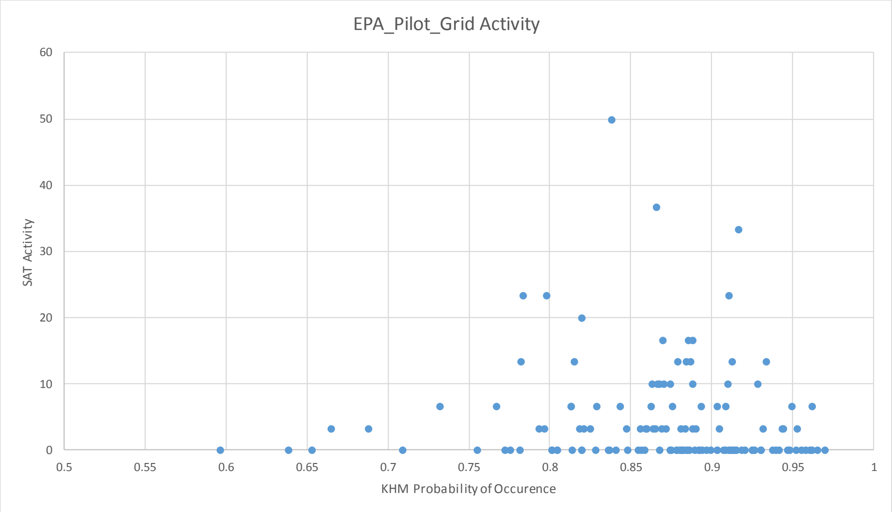


Predicted occurrence vs SAT Activity









Summary & Recommendations

- SAT activity is almost always zero at sites below the 0.65 probability of occurrence threshold. This suggests the model is reasonably accurate in identifying areas that are unsuitable or of low suitability.
- Above the 0.65 threshold, activity scores are highly variable. This pattern is common in ecological studies that try to relate animal abundance to predicted habitat suitability. The pattern can arise if:
  - a) a large proportion of moderate to high quality habitat is unoccupied at the time of survey (i.e. due to past disturbance events such as fire, drought, predation or disease); and/or
  - b) methods to assess activity are limited and only detect a small portion of sites that are actually occupied.

Summary & Recommendations

- The BRT model presented is useful for predicting “potential” koala habitat (areas of moderate to high suitability) but has no power to distinguish habitats that are occupied from those that are unoccupied by koalas within areas of “potential habitat”.
- It is possible model the distributions of different classes of habitat such as core refugia, primary, secondary, marginal and non-habitat in a single multinomial model. However, such a model would require a detailed knowledge of local koala habitat preferences and the vegetation at those sites.
- Future modelling should try to take account of regional variations in food tree preferences (where such knowledge exists).
- Even with improved models, the occupancy of habitats may still be limited by social factors, past disturbance events or dispersal limitations.
- Since most of these factors can not be incorporated into models - ground surveys will remain the only reliable option for identifying sites that are actually occupied.