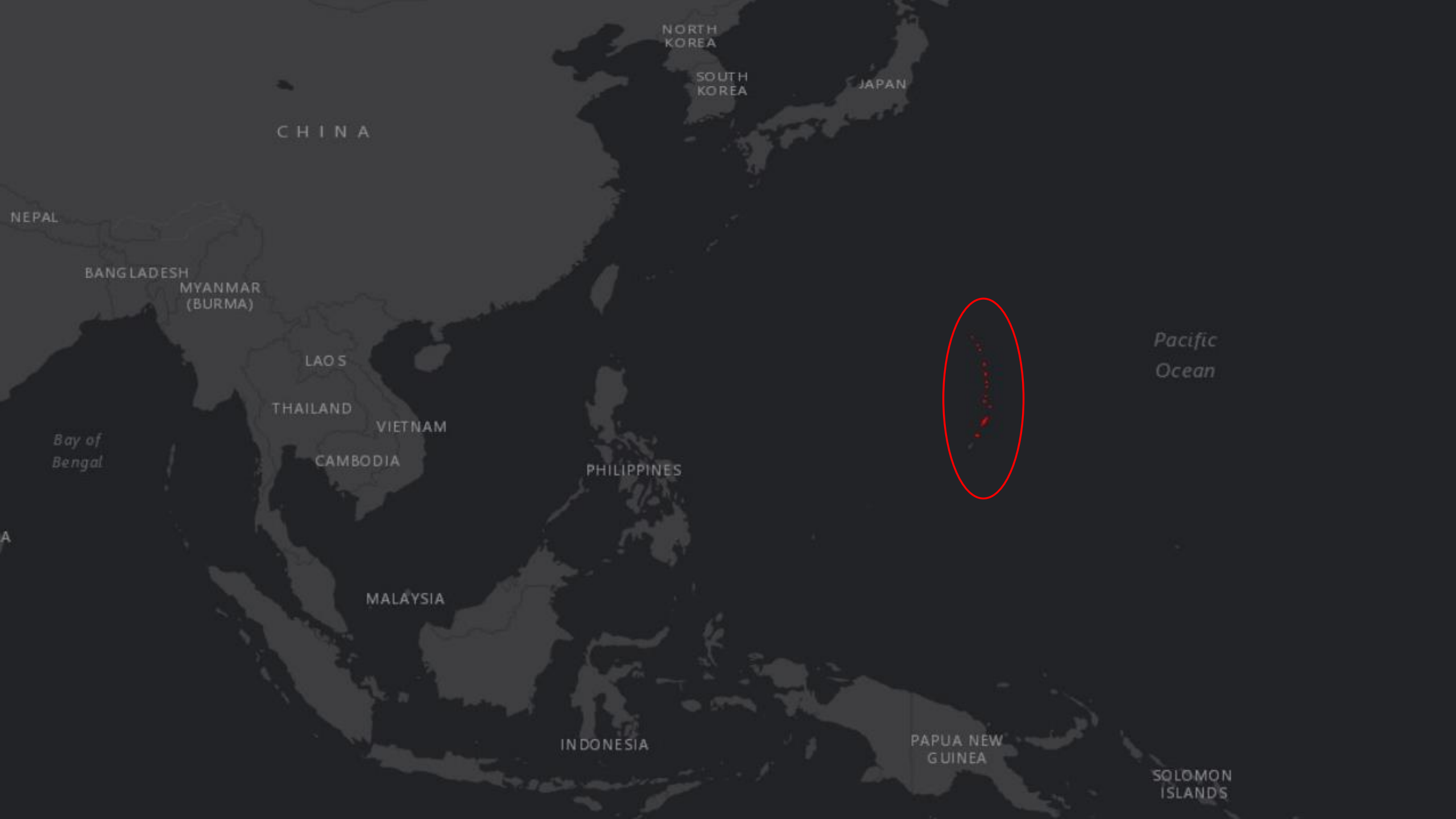
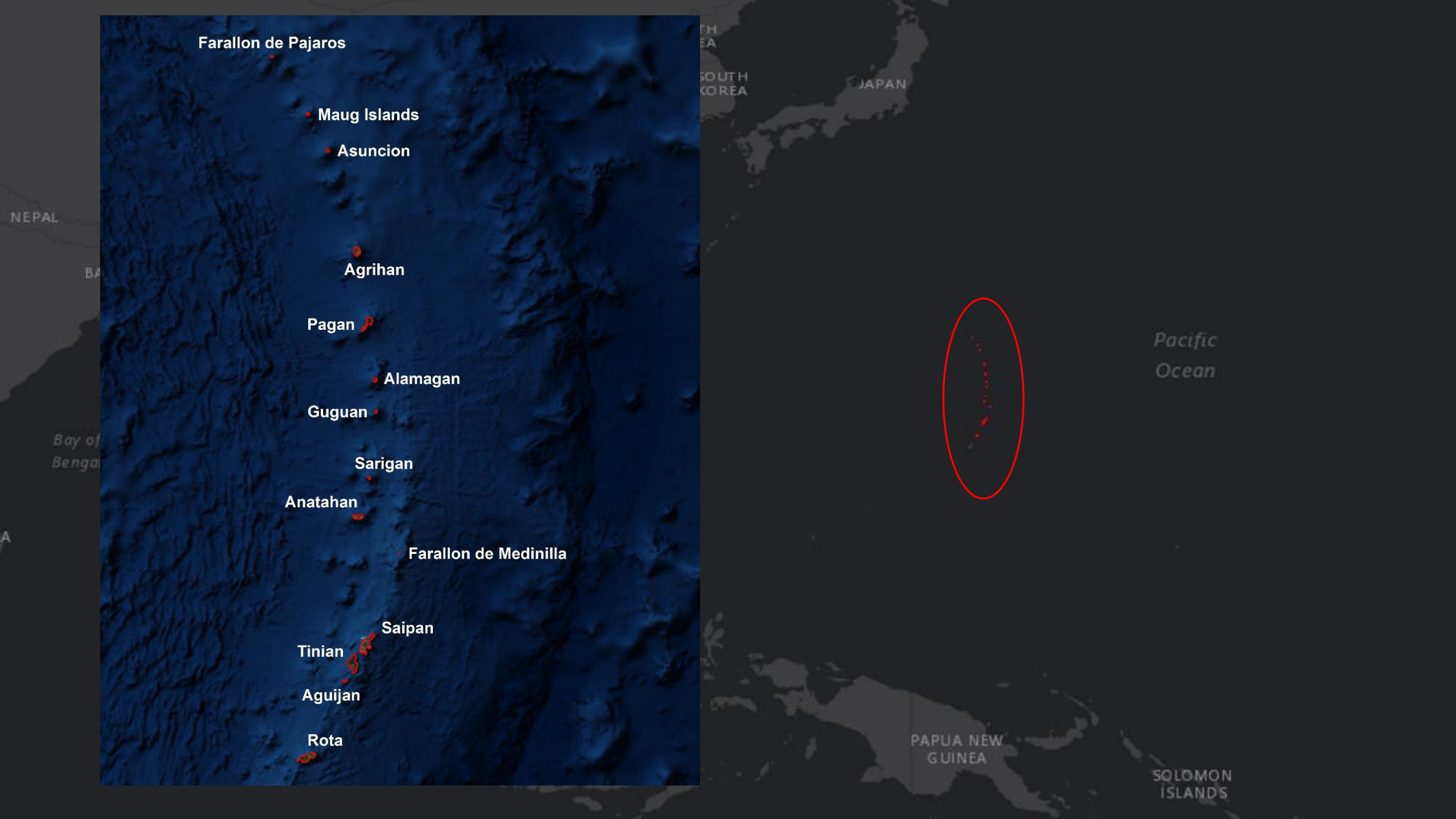
# PREDICTING DISTRIBUTIONS OF ENDEMIC BIRD SPECIES IN THE NORTHERN MARIANA ISLANDS

BRAD EICHELBERGER

DIVISION OF FISH AND WILDLIFE

COMMONWEALTH OF THE NORTHERN MARIANA ISLANDS

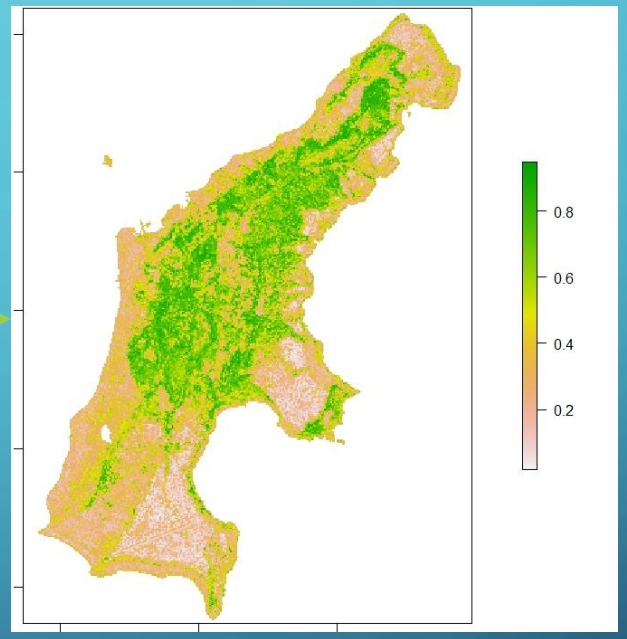

NORTH KOREA
SOUTH KOREA
JAPAN
CHINA
NEPAL
BANGLADESH
MYANMAR (BURMA)
LAOS
THAILAND
VIETNAM
CAMBODIA
Bay of Bengal
PHILIPPINES
MALAYSIA
INDONESIA
PAPUA NEW GUINEA
SOLOMON ISLANDS
Pacific Ocean


Farallon de Pajaros
Maug Islands
Asuncion
Agrihan
Pagan
Alamagan
Guguan
Sarigan
Anatahan
Farallon de Medinilla
Saipan
Tinian
Aguijan
Rota
NEPAL
SOUTH KOREA
JAPAN
Pacific Ocean
PAPUA NEW GUINEA
SOLOMON ISLANDS

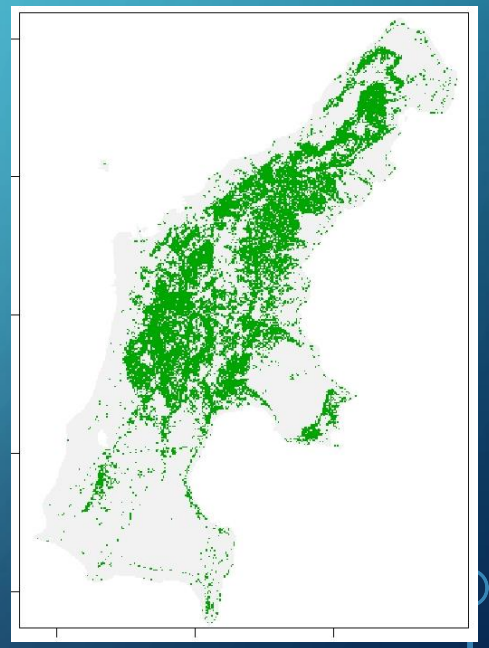
# BACKGROUND

- 12 endemic bird species
  - 6 federally listed
- Threats: habitat loss, development, brown tree snake
- Critical to understand habitat requirements
  - Focus future surveys
  - Population estimates
  - Protect or restore habitat

# SPECIES DISTRIBUTION MODELS


Threshold

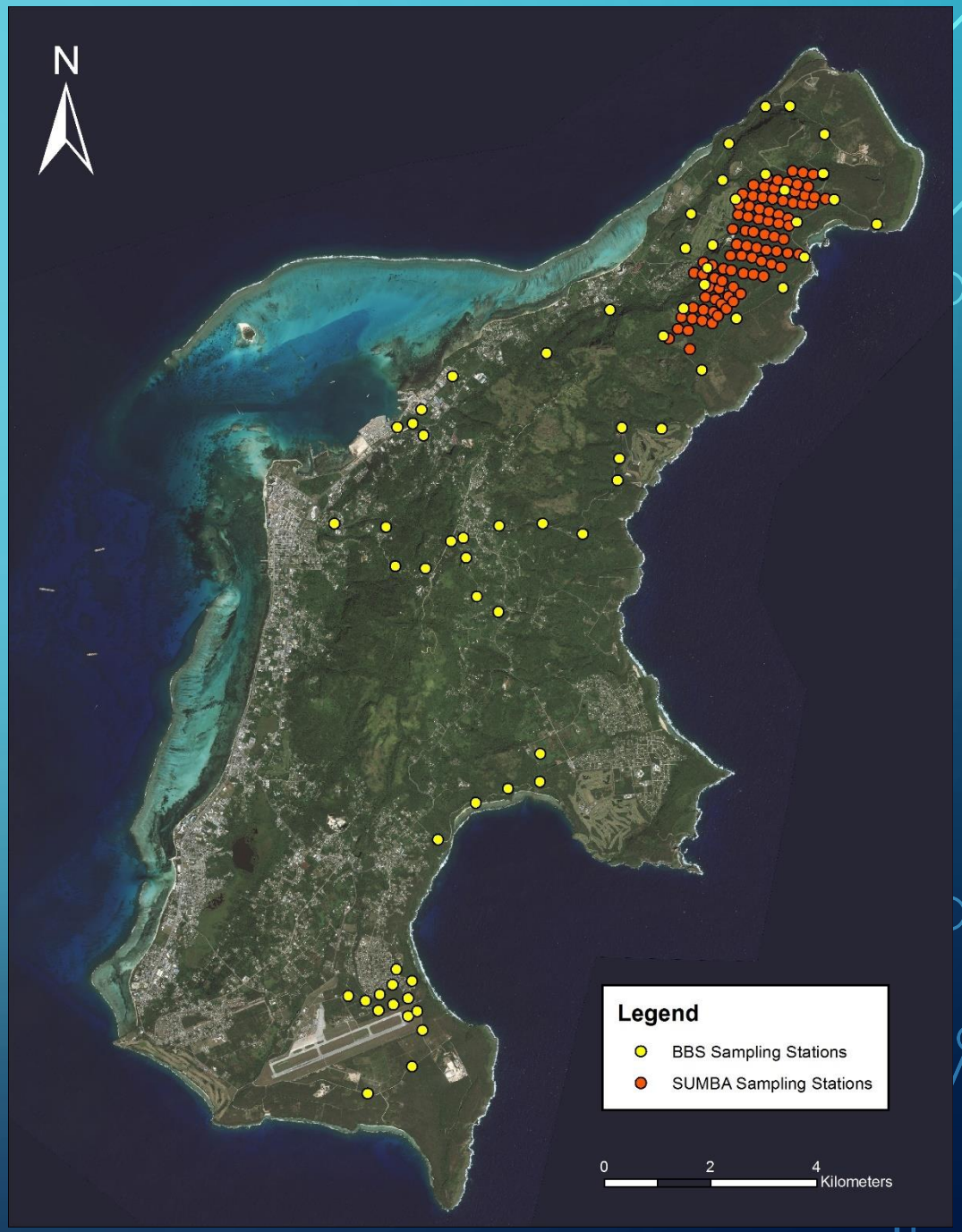


# MARIANA FRUIT DOVE

- IUCN - Globally Endangered
- Frugivore
- Forest - top canopy
- Wide range

# SAMPLE DATA

- BBS data (n = 64)
  - Quarterly 2016-2017
- SUMBA (n = 88)
  - 2017
- Only occurrences within 25-m radius

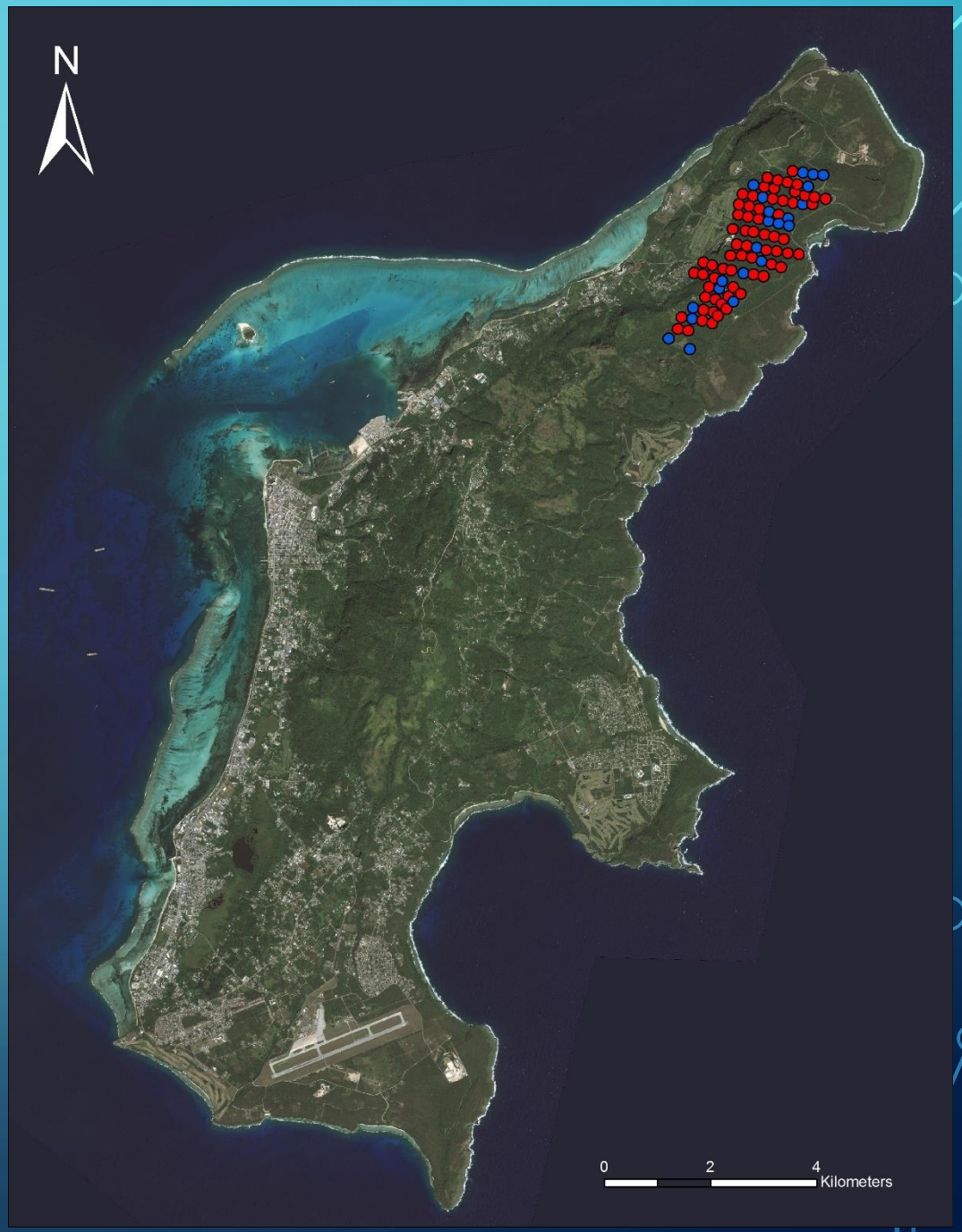


# SAMPLE DATA

- BBS data (n = 64)
  - Quarterly 2016-2017
  - Presence = 26
  - Absence = 38

# SAMPLE DATA

- SUMBA (n = 88)
  - 2017
  - Presence = 23
  - Absence = 65



# PREDICTOR VARIABLES

- Habitat composition
  - % cover around 100-$m^2$ area – USFWS vegetation map (2016) 2-m resolution (21 classes)
- Topographic variables
  - Elevation (NOAA 5-m resolution)
  - % slope (NOAA 5-m resolution)
  - Terrain roughness - topographic index
    - (elev.(mean) – elev.(min)) / (elev.(max) – elev.(min))
- Enhanced Vegetation Index (EVI)
  - Sentinel2 satellite imagery 10-m resolution
  - Wet season (10/25/2016 - 12/15/2016)
  - Dry season (5/13/2017 - 6/15/2017)

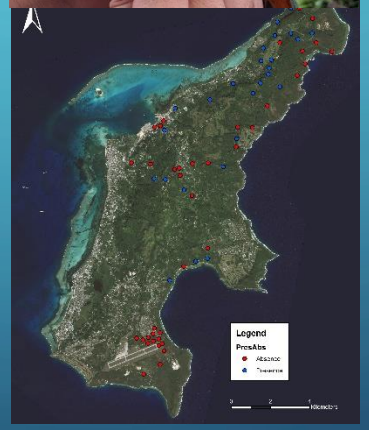
# PREDICTOR VARIABLES

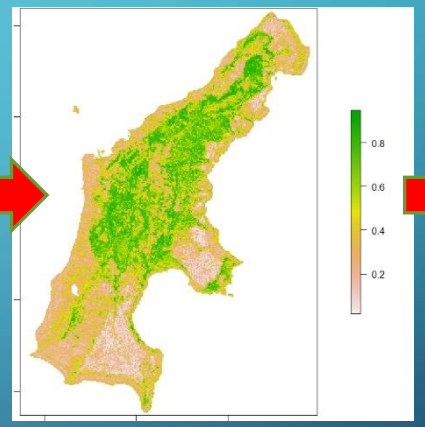
- Habitat composition
  - % cover around 100-$m^2$ area – USFWS vegetation map (2016) 2-m resolution (21 classes)
- Topographic variables
  - Elevation (NOAA 5-m resolution)
  - % slope (NOAA 5-m resolution)
  - Terrain roughness - topographic index
    - (elev.(mean) – elev.(min)) / (elev.(max) – elev.(min))
- Enhanced Vegetation Index (EVI)
  - Sentinel2 satellite imagery 10-m resolution
  - Wet season (10/25/2016 - 12/15/2016)
  - Dry season (5/13/2017 - 6/15/2017)

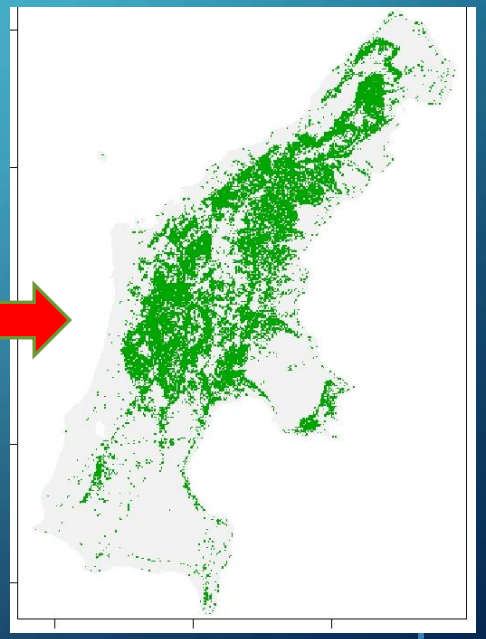
All variables are resampled to 50-meter resolution to match sample uncertainty!

# SPECIES DISTRIBUTION MODELS

- Logistic Regression (stepwise)
- Generalized Additive Models
  - Stepwise
  - Smoother = 3
- Maximum Entropy
- Random Forest
- Boosted Regression Trees
  - Tree complexity = 3, learning rate = 0.0005
- Ensemble models



# VALIDATION METHODS

- Threshold – probabilities to habitat (maximize kappa coefficient)
- Measures of accuracy
  - Sensitivity - rate of true positives
  - Specificity - rate of true negatives
  - Area Under Curve (AUC) - model performance
  - True Skill Statistic (TSS) - similar to kappa (-1.0 – 1.0)
- Three levels
  - Internal - (BBS data)
  - Cross validation - (BBS data) 80% for training, 20% for validation
  - External - (SUMBA data)

# VALIDATION METHODS

- Threshold – probabilities to habitat (maximize kappa coefficient)
- Measures of accuracy
  - Sensitivity - rate of true positives
  - Specificity - rate of true negatives
  - Area Under Curve (AUC) - model performance
  - True Skill Statistic (TSS) - similar to kappa (-1.0 – 1.0)
- Three levels
  - Internal - (BBS data)
  - Cross validation - (BBS data) 80% for training, 20% for validation
  - External - (SUMBA data)

Closer to 1.0, the better the model is!!!

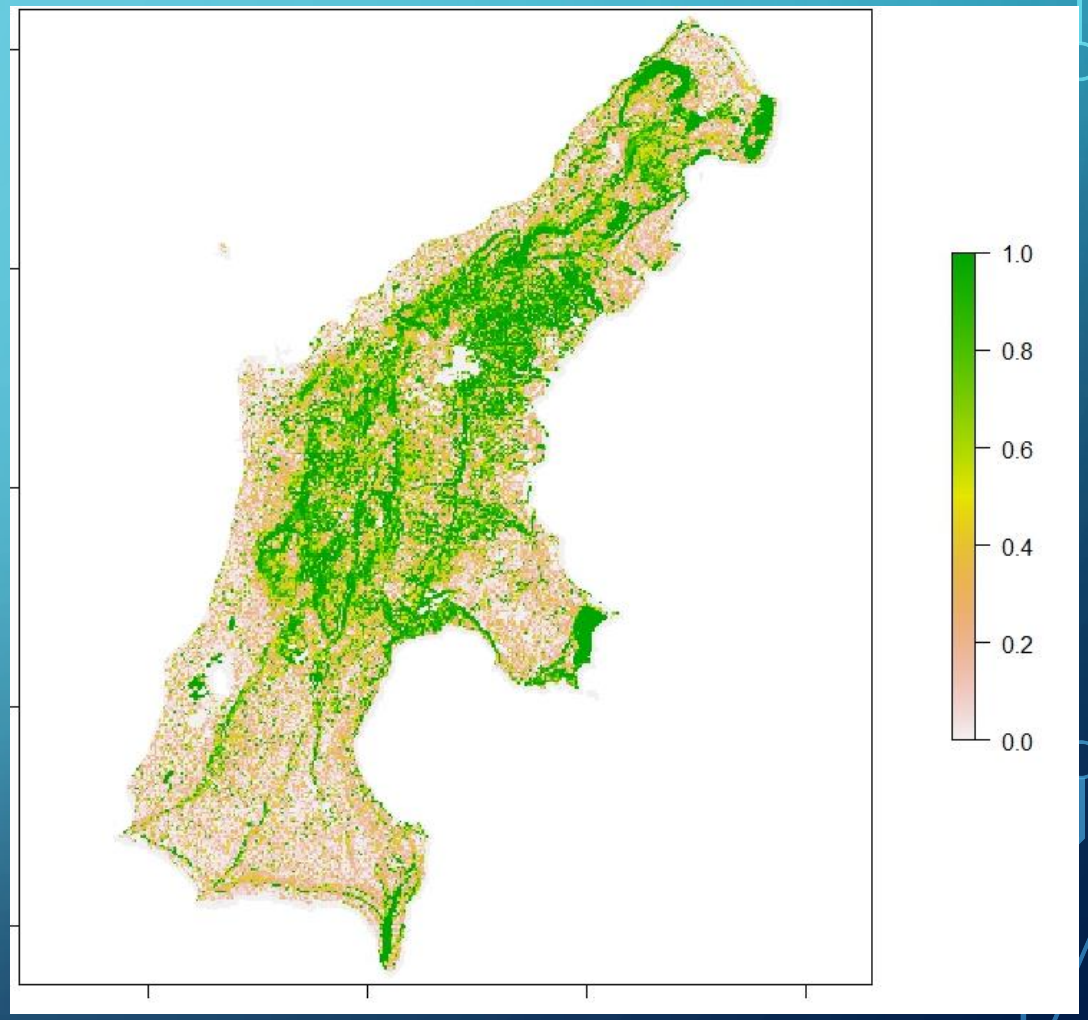
# LOGISTIC REGRESSION

| | Internal | Cross-Val | External |
|---|---|---|---|
| **Sensitivity** | **0.85** | **0.81** | **0.68** |
| **Specificity** | **0.84** | **0.71** | **0.35** |
| **AUC** | **0.9** | **0.81** | **--** |
| **TSS** | **0.69** | **0.52** | **0.03** |

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -15.81857 | 5.18652 | -3050 | 0.002289 ** |
| SLOPE | 0.15200 | 0.04578 | 3.321 | 0.000898 *** |
| HIBISCUS_T | 0.01363 | 4.29614 | 0.003 | 0.997469 |
| ROUGHNESS | 26.00893 | 9.57656 | 2.716 | 0.006610 ** |
| OPEN_WATER | -0.42951 | 5355.12069 | 0.000 | 0.999936 |
| CASUARINA | -0.09244 | 31.37529 | -0.003 | 0.997649 |
| NATIVE_LIM | 0.01549 | 1.96865 | 0.008 | 0.993722 |

$D^2 = 43.2$

Threshold = 0.36



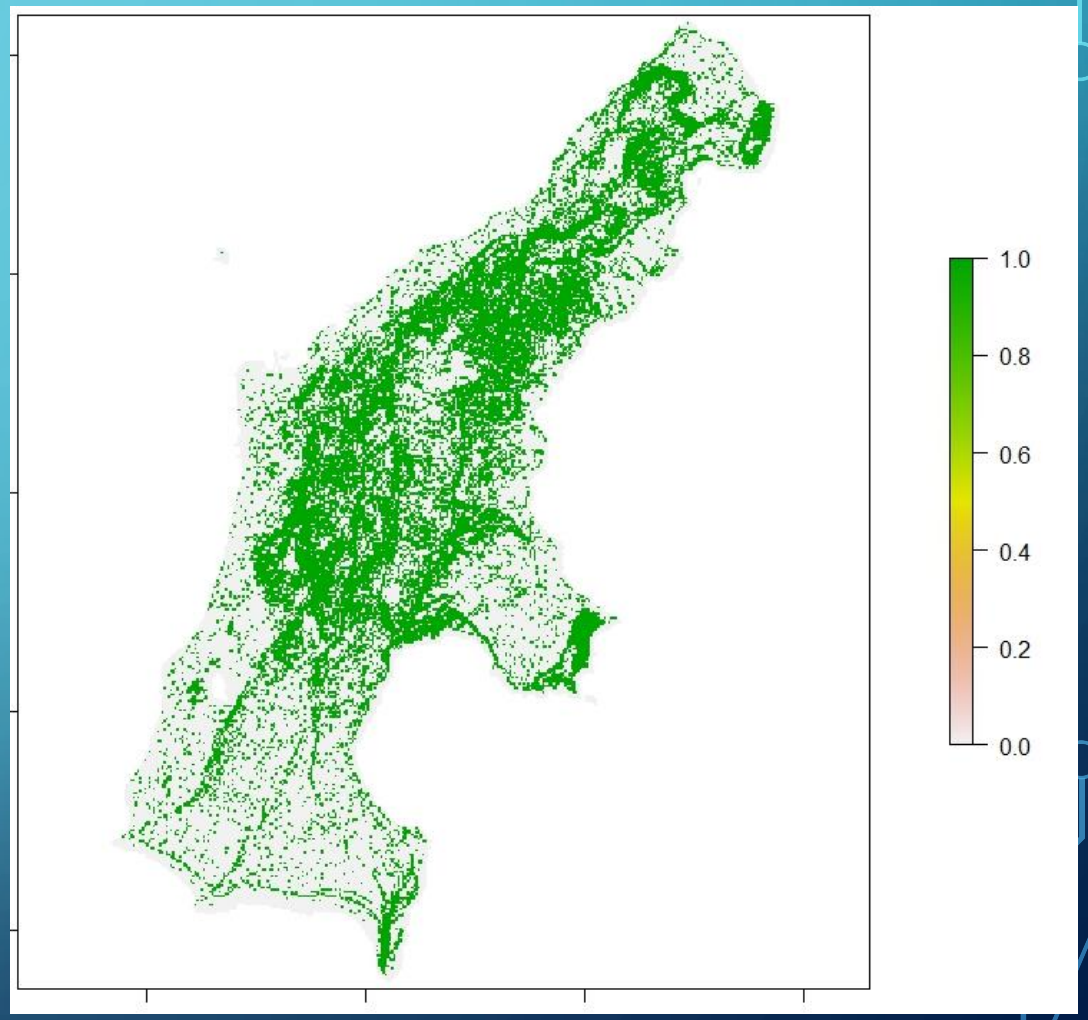
# LOGISTIC REGRESSION

| | Internal | Cross-Val | External |
|---|---|---|---|
| **Sensitivity** | **0.85** | **0.81** | **0.68** |
| **Specificity** | **0.84** | **0.71** | **0.35** |
| **AUC** | **0.9** | **0.81** | **--** |
| **TSS** | **0.69** | **0.52** | **0.03** |

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -15.81857 | 5.18652 | -3050 | 0.002289 ** |
| SLOPE | 0.15200 | 0.04578 | 3.321 | 0.000898 *** |
| HIBISCUS_T | 0.01363 | 4.29614 | 0.003 | 0.997469 |
| ROUGHNESS | 26.00893 | 9.57656 | 2.716 | 0.006610 ** |
| OPEN_WATER | -0.42951 | 5355.12069 | 0.000 | 0.999936 |
| CASUARINA | -0.09244 | 31.37529 | -0.003 | 0.997649 |
| NATIVE_LIM | 0.01549 | 1.96865 | 0.008 | 0.993722 |

$D^2 = 43.2$

Threshold = 0.36



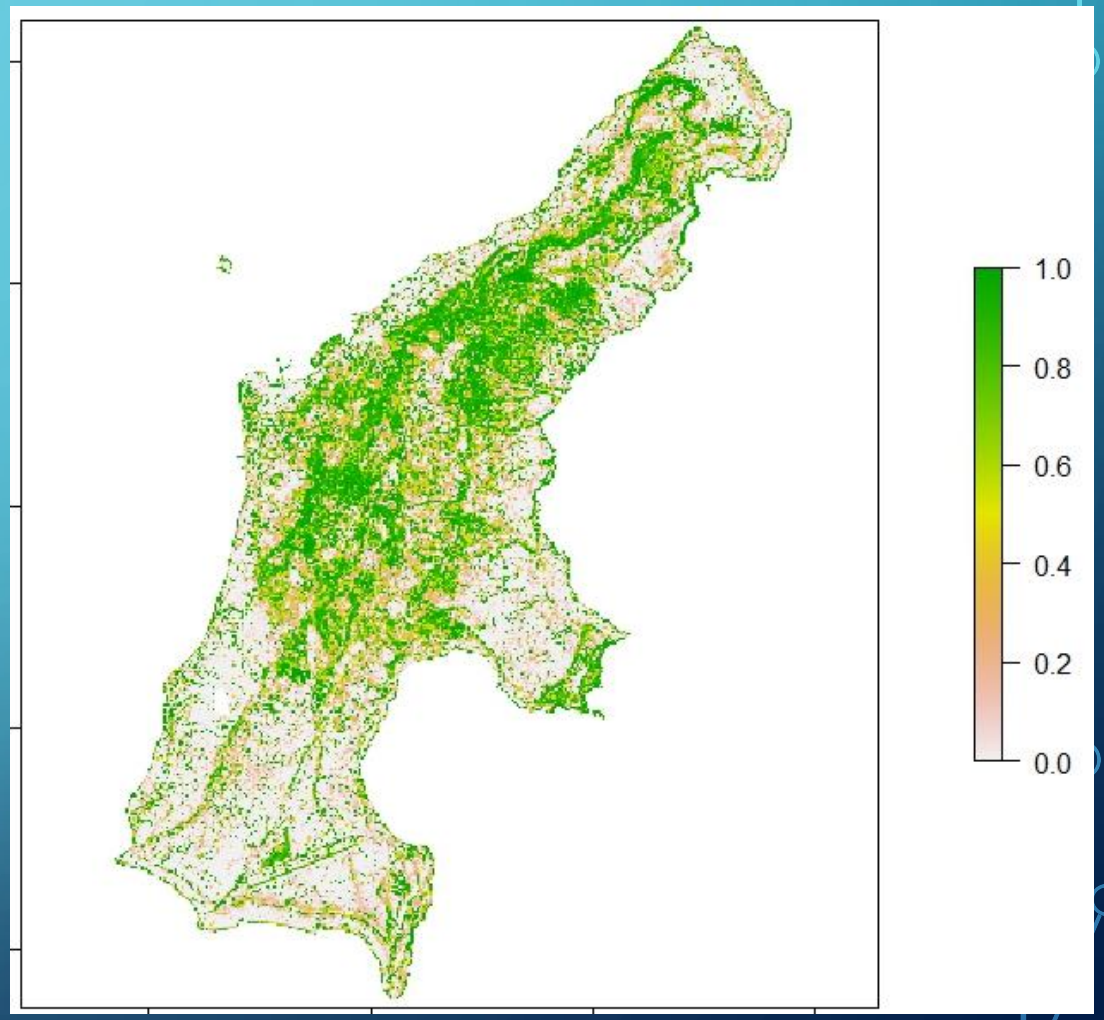
# GENERALIZED ADDITIVE MODEL

| | Internal | Cross-Val | External |
|---|---|---|---|
| **Sensitivity** | **1.0** | **0.77** | **0.55** |
| **Specificity** | **0.87** | **0.71** | **0.35** |
| **AUC** | **0.97** | **0.78** | **--** |
| **TSS** | **0.87** | **0.48** | **-0.11** |

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| s(SLOPE, 3) | 1 | 11.0797 | 11.0797 | 22.3781 | 1.811e-05 *** |
| s(LEUCAENA, 3) | 1 | 0.0324 | 0.0324 | 0.0654 | 0.799206 |
| s(EVI_Wet, 3) | 1 | 5.8405 | 5.8405 | 11.7962 | 0.001187 ** |
| s(ROUGHNESS, 3) | 1 | 1.6471 | 1.6471 | 3.3268 | 0.074021 . |

$D^2 = 60.7$

Threshold = 0.35



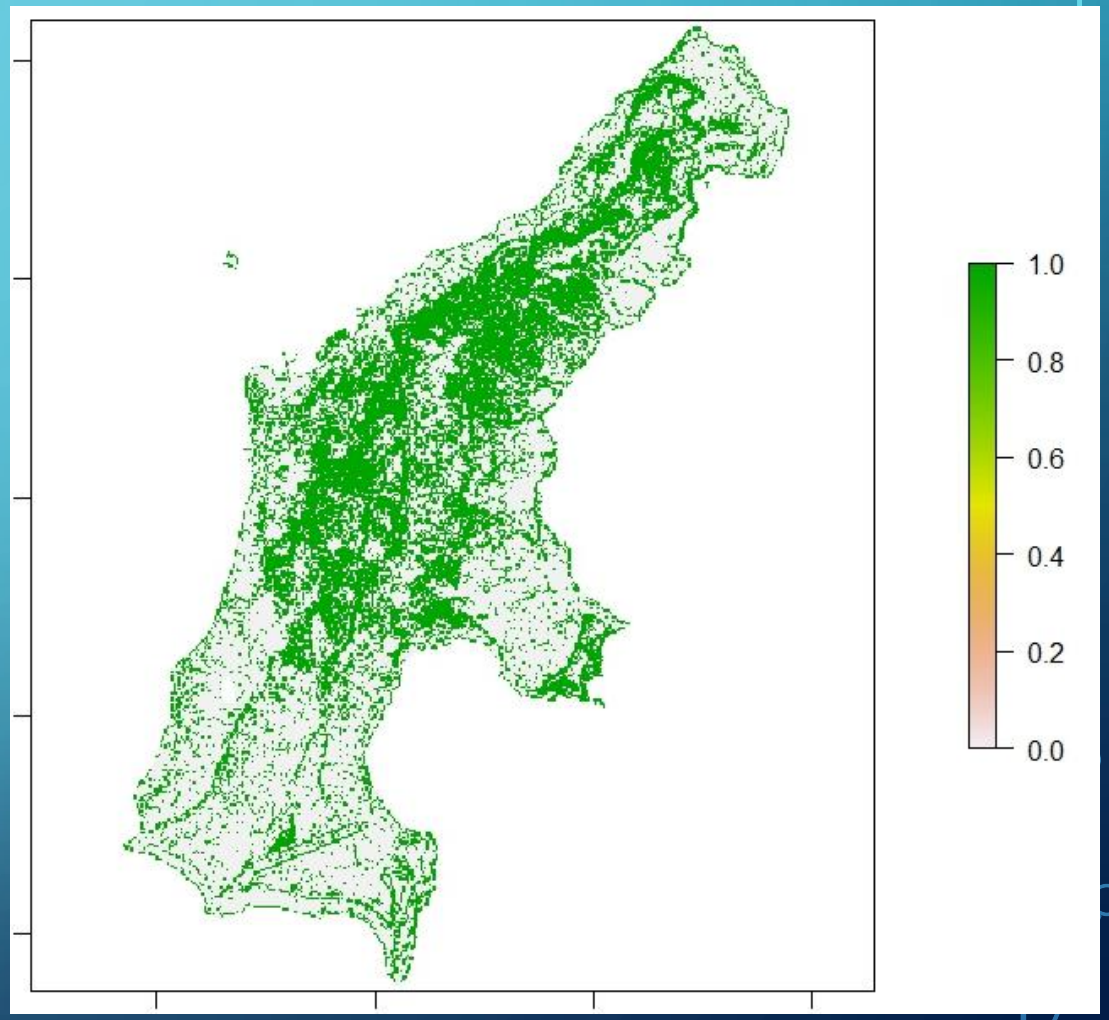
# GENERALIZED ADDITIVE MODEL

| | Internal | Cross-Val | External |
|---|---|---|---|
| **Sensitivity** | **1.0** | **0.77** | **0.55** |
| **Specificity** | **0.87** | **0.71** | **0.35** |
| **AUC** | **0.97** | **0.78** | **--** |
| **TSS** | **0.87** | **0.48** | **-0.11** |

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| s(SLOPE, 3) | 1 | 11.0797 | 11.0797 | 22.3781 | 1.811e-05 *** |
| s(LEUCAENA, 3) | 1 | 0.0324 | 0.0324 | 0.0654 | 0.799206 |
| s(EVI_Wet, 3) | 1 | 5.8405 | 5.8405 | 11.7962 | 0.001187 ** |
| s(ROUGHNESS, 3) | 1 | 1.6471 | 1.6471 | 3.3268 | 0.074021 . |

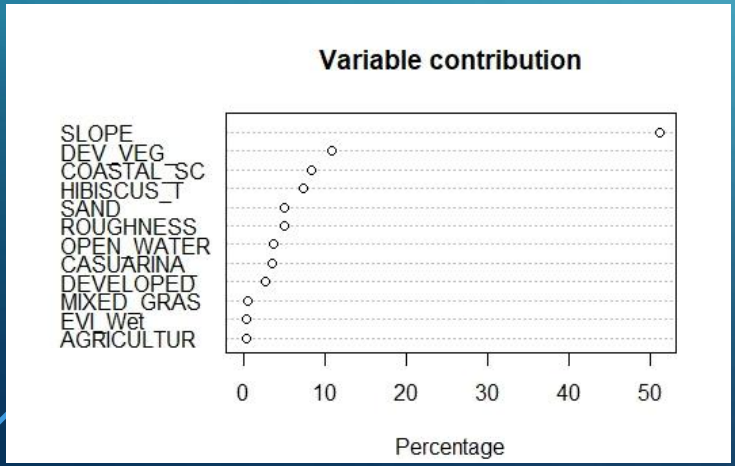
$D^2 = 60.7$

Threshold = 0.35

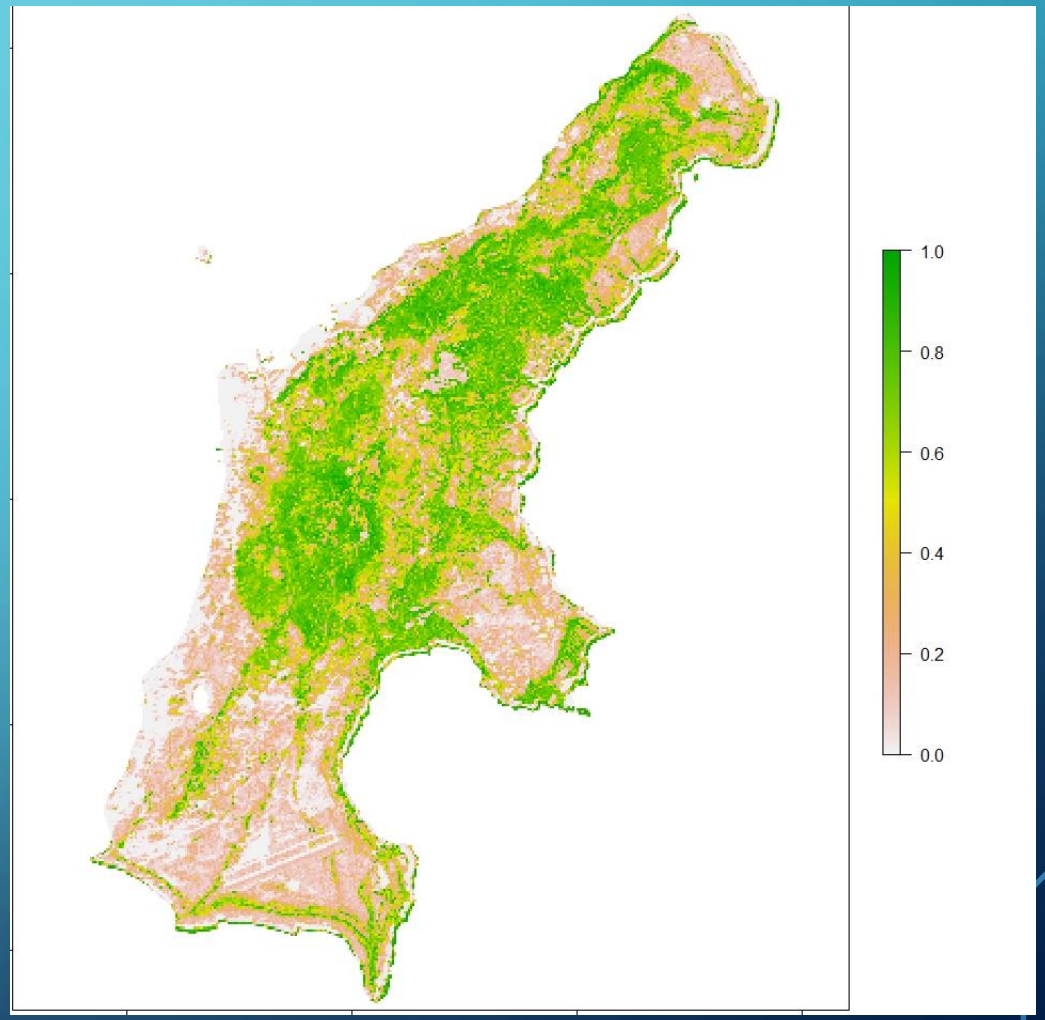


# MAXIMUM ENTROPY

| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.81 | 0.46 | 0.77 |
| Specificity | 0.97 | 1.0 | 0.18 |
| AUC | 0.96 | 0.71 | -- |
| TSS | 0.78 | 0.46 | -0.05 |

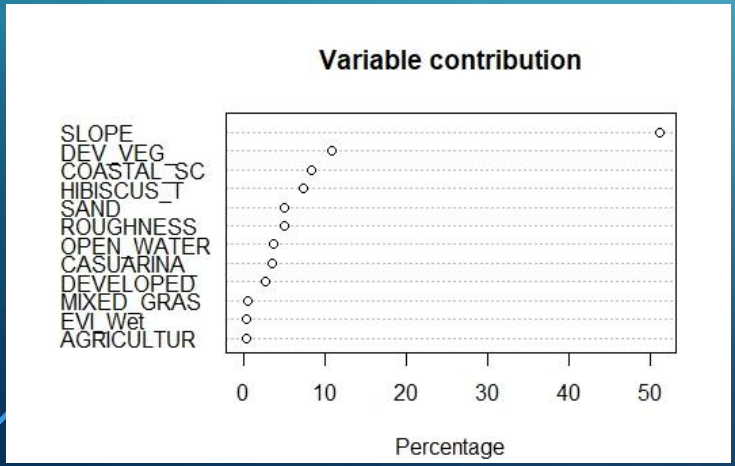



Threshold = 0.60

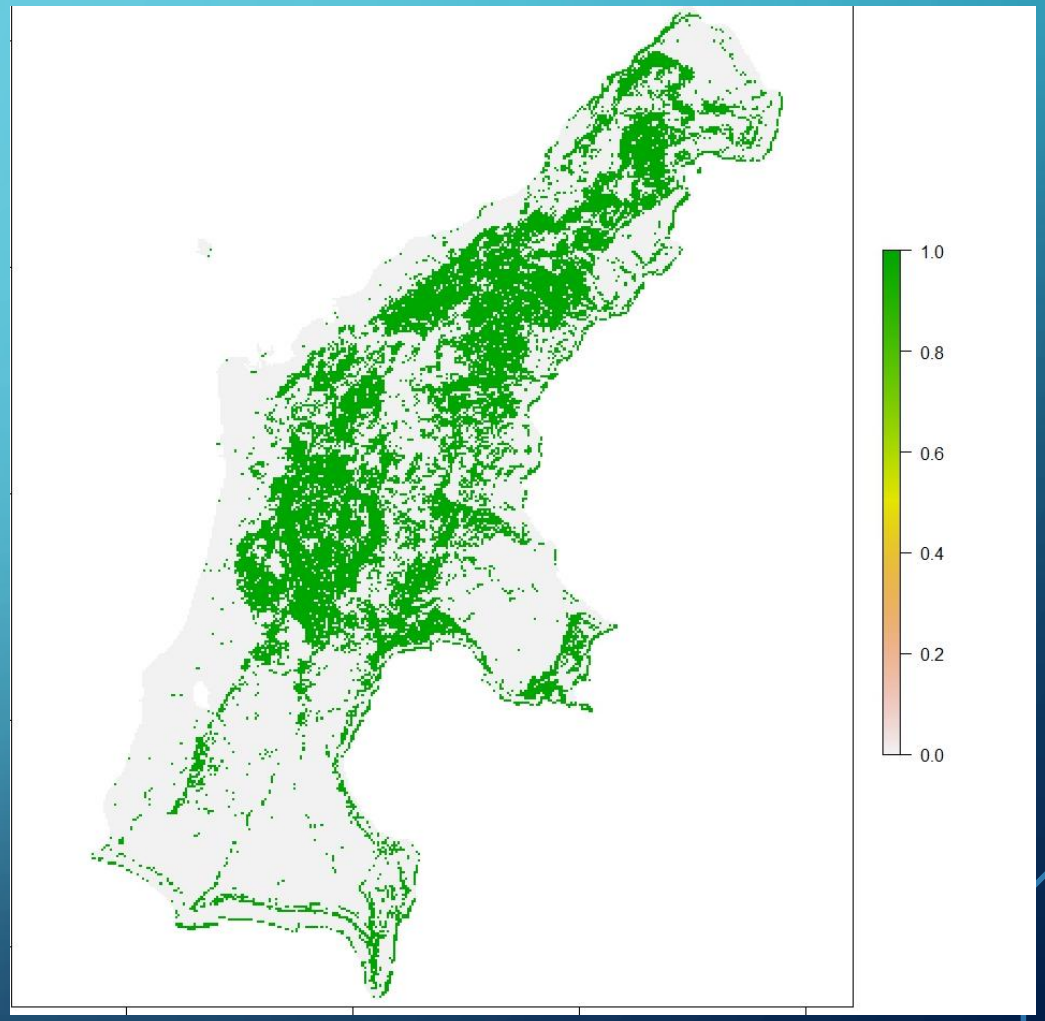


# MAXIMUM ENTROPY

| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.81 | 0.46 | 0.77 |
| Specificity | 0.97 | 1.0 | 0.18 |
| AUC | 0.96 | 0.71 | -- |
| TSS | 0.78 | 0.46 | -0.05 |



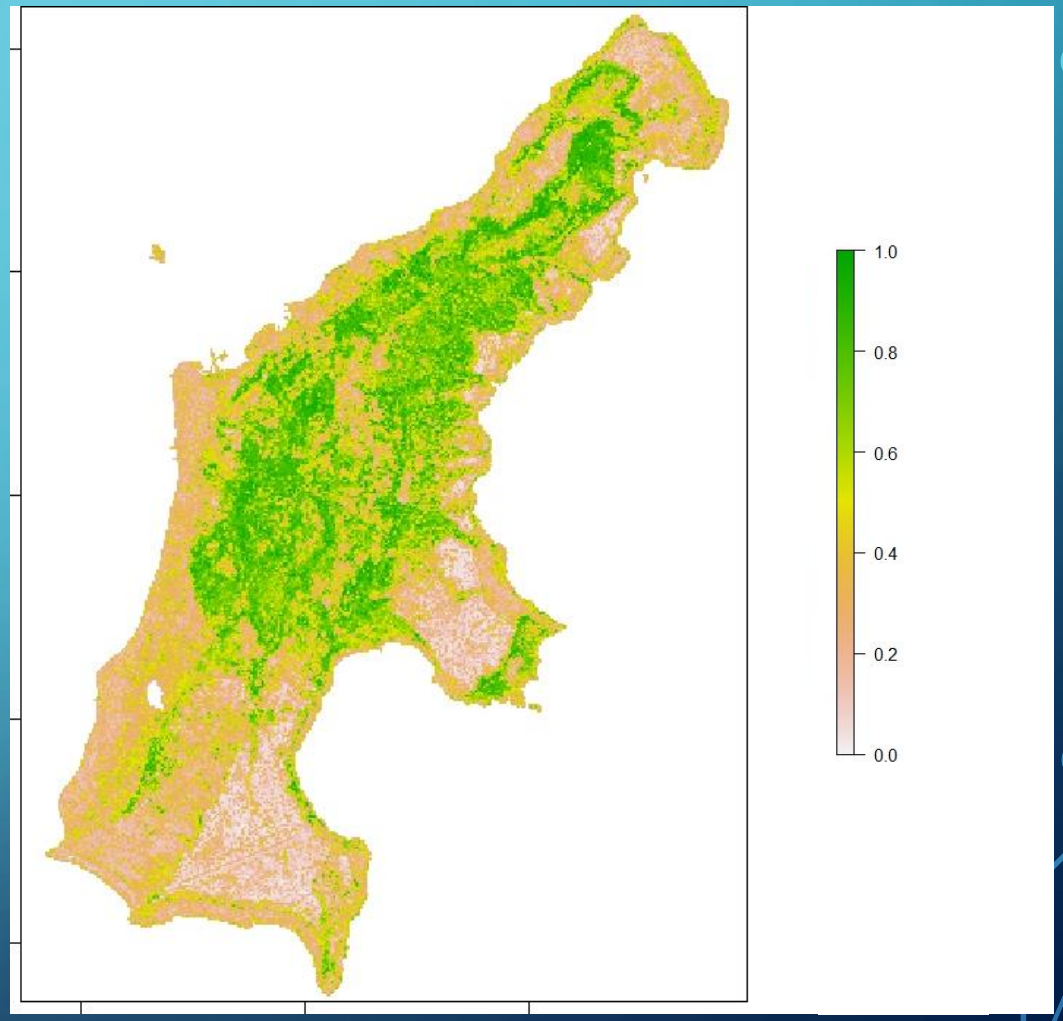

Threshold = 0.60



# RANDOM FOREST

| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.5 | 0.53 | 0.68 |
| Specificity | 0.92 | 0.81 | 0.40 |
| AUC | 0.74 | 0.75 | -- |
| TSS | 0.42 | 0.35 | 0.08 |

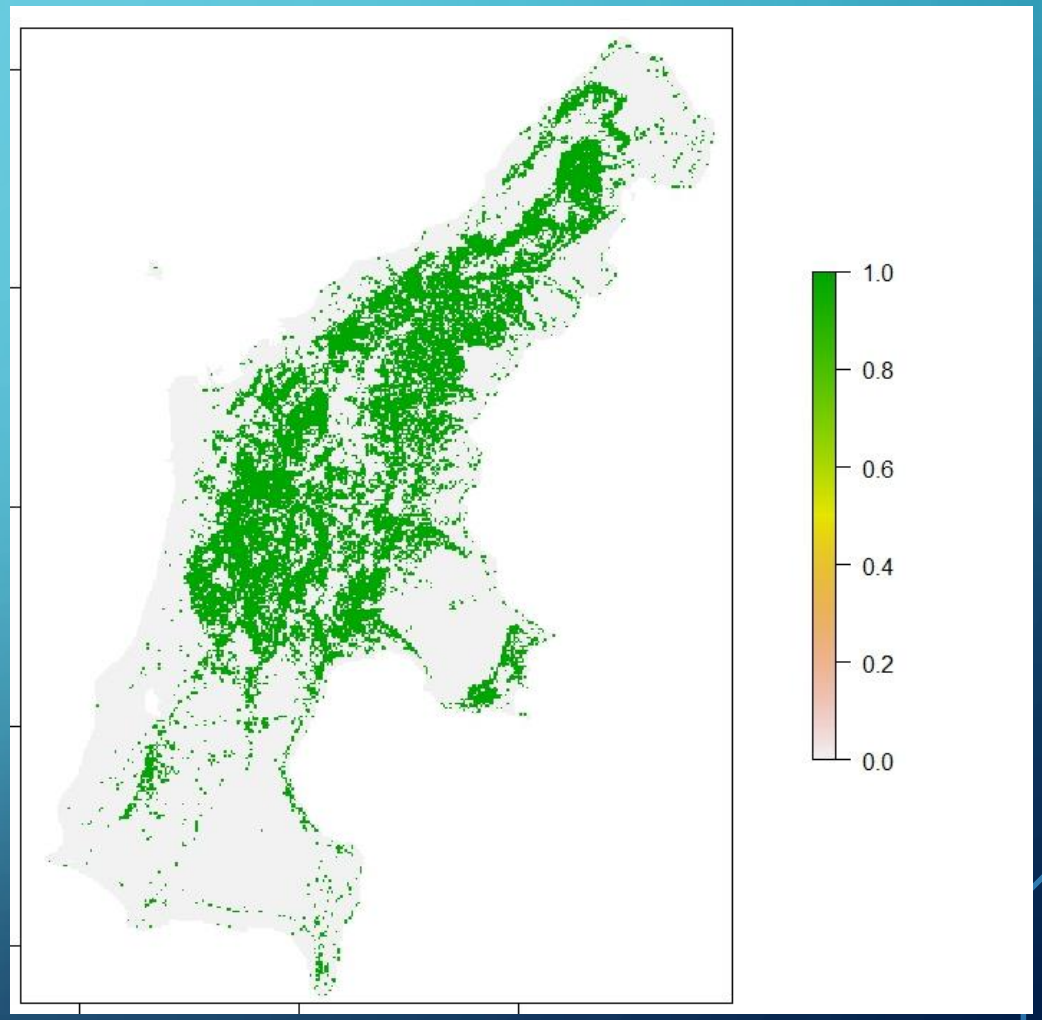
Threshold = 0.65



# RANDOM FOREST

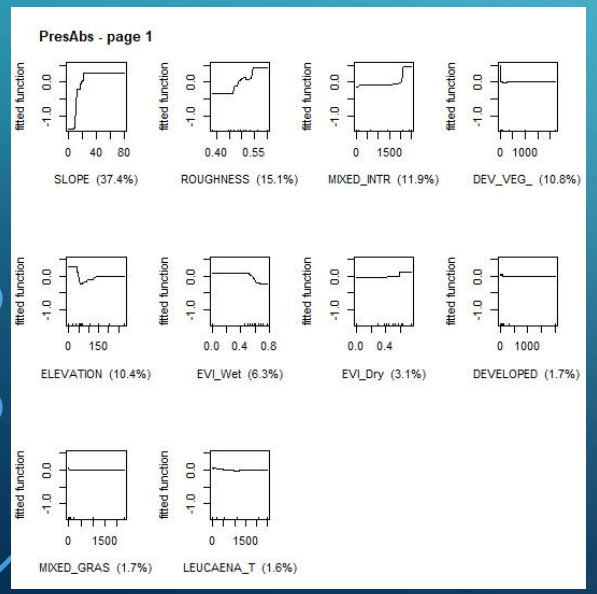
| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.5 | 0.53 | 0.68 |
| Specificity | 0.92 | 0.81 | 0.40 |
| AUC | 0.74 | 0.75 | -- |
| TSS | 0.42 | 0.35 | 0.08 |

Threshold = 0.65

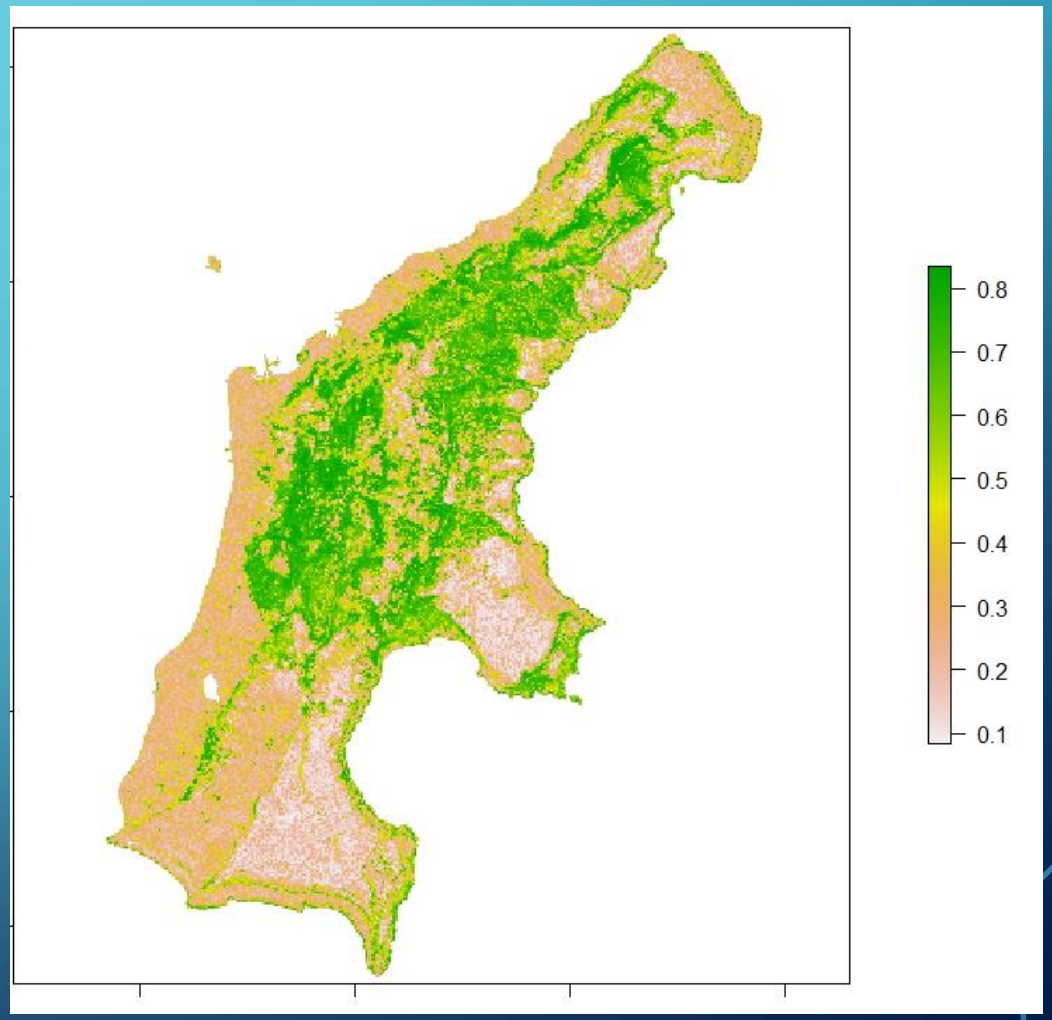


# BOOSTED REGRESSION TREES

| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.96 | 0.61 | 0.78 |
| Specificity | 0.94 | 0.78 | 0.28 |
| AUC | 0.99 | 0.75 | -- |
| TSS | 0.91 | 0.41 | 0.05 |

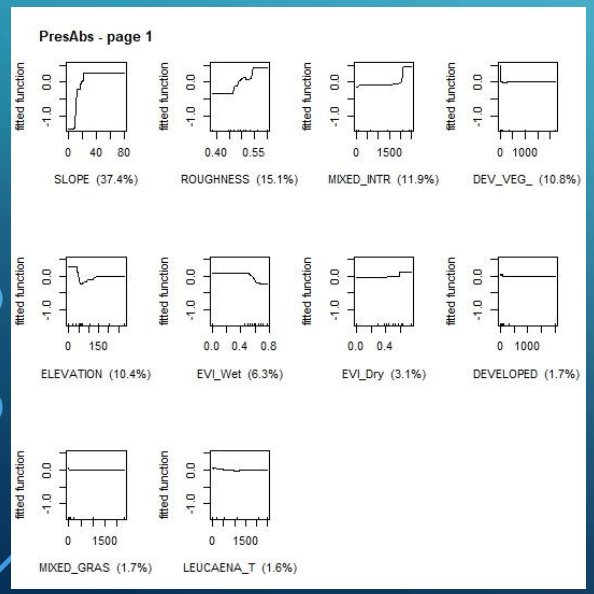



Threshold = 0.41

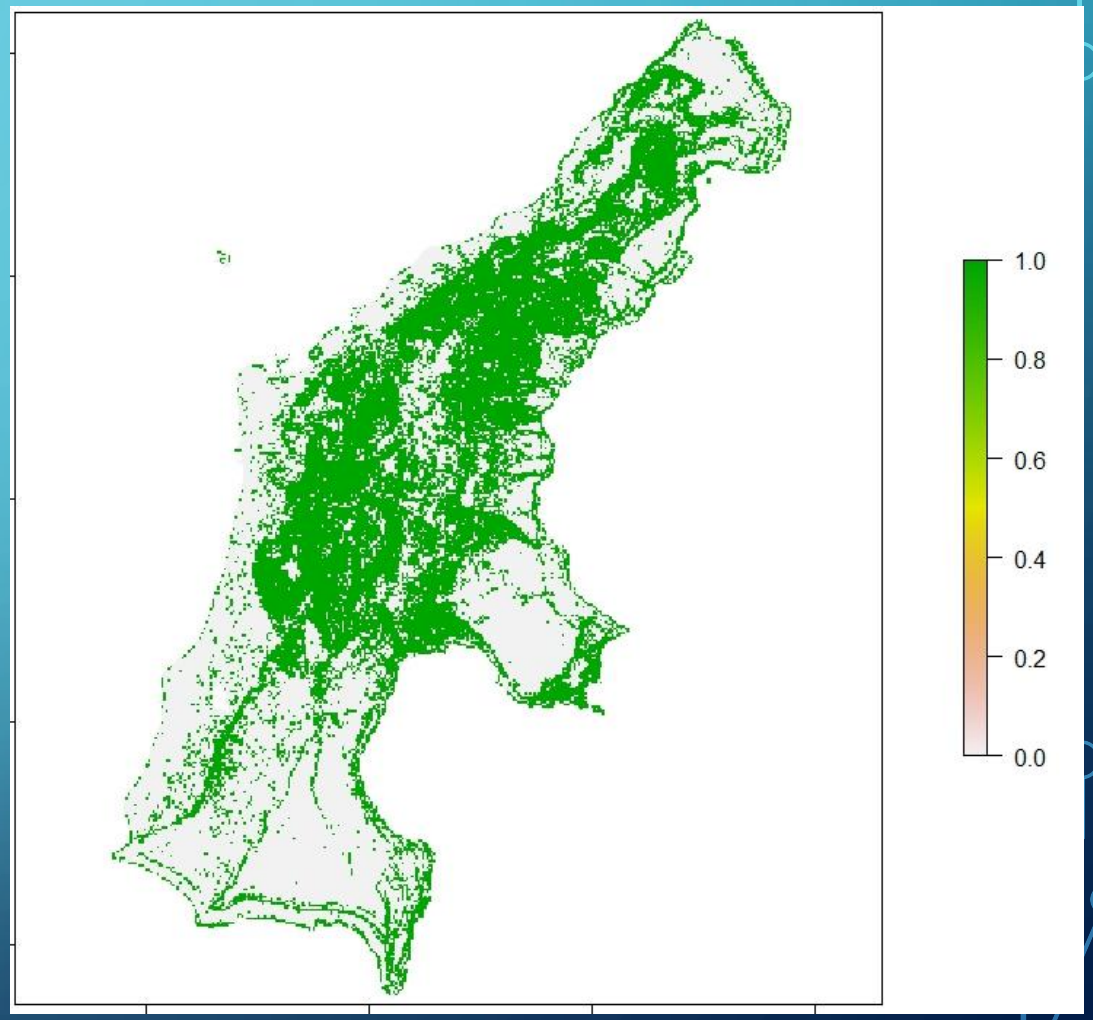


# BOOSTED REGRESSION TREES

| | Internal | Cross-Val | External |
|---|---|---|---|
| Sensitivity | 0.96 | 0.61 | 0.78 |
| Specificity | 0.94 | 0.78 | 0.28 |
| AUC | 0.99 | 0.75 | -- |
| TSS | 0.91 | 0.41 | 0.05 |

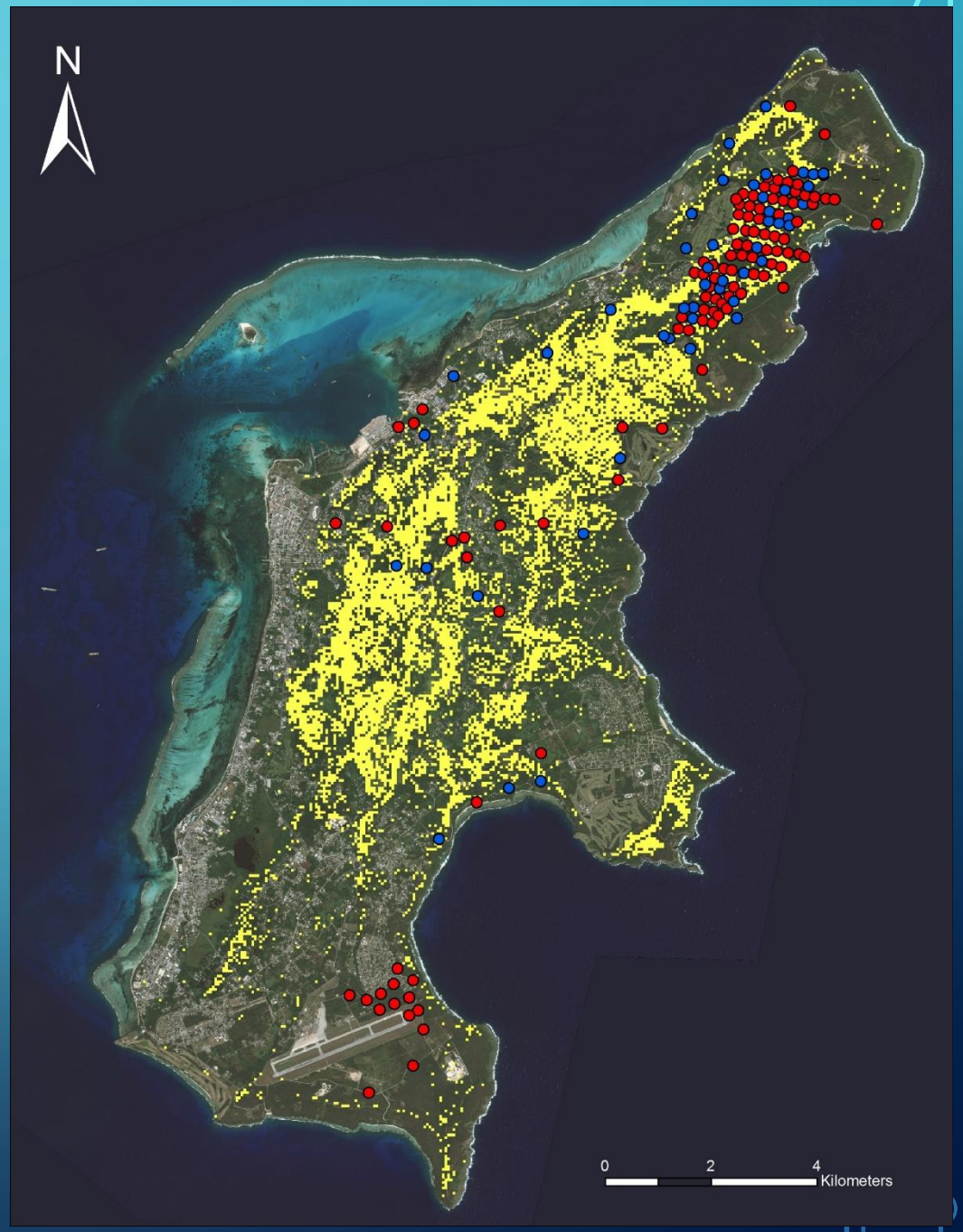



Threshold = 0.41



# ENSEMBLE MAPS

- All models
  - 2,554 ha. (22% of Saipan)
  - Sensitivity = 0.40
  - Specificity = 0.55
  - TSS = -0.05
- Drop RF for 4 model ensemble
  - 3,025 ha. (26% of Saipan)
  - Sensitivity = 0.45
  - Specificity = 0.50
  - TSS = -0.05



# CONCLUSIONS

- Models seem to overfit data however….
  - Data distribution, collection, and quality are important
  - Low sample size
  - Species has large area of coverage, SUMBA one time sample
- Slope, roughness, and EVI reoccurring variables – habitat complexity or sampling bias?

# FUTURE STEPS

- Improved data collection - Saipan wide surveys
- Hierarchical models incorporating abundance for detectability over time
- Habitat configuration for conservation planning

# ACKNOWLEDGEMENTS


- Funding source - USFWS WSPR

- People - Steve Mullin, Lainie Berry, Jay Camacho, Frances Sablan, Chris Murray, Jill Liske-Clark, Robert Ulloa, Marlyn Naputi, Lorraine Reyes, Tyler Willsey