Using Random Forest Regression to Interpolate ICESat-2 Elevation Data

Ling Wo Leo Kan #5505801

Supervisor 1: Hugo Ledoux

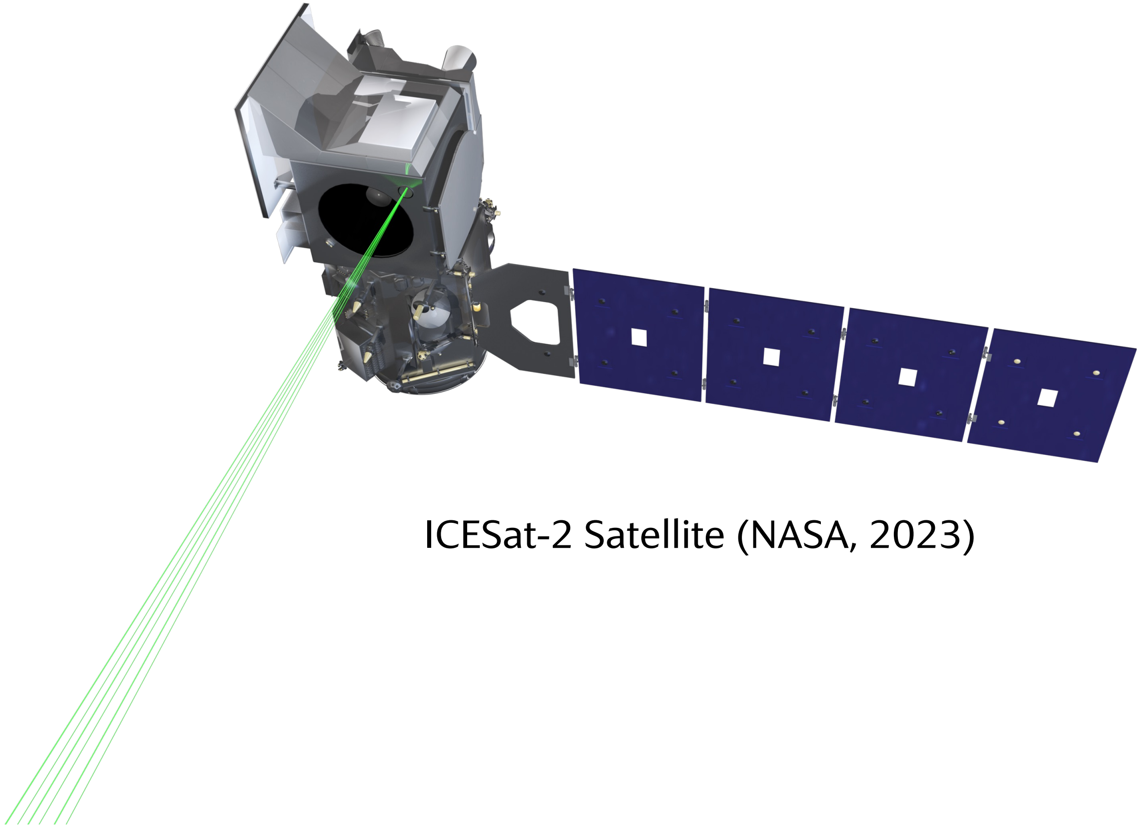
Supervisor 2: Maarten Pronk

Co-reader: Azarakhsh Rafiee

TU Delft

Contents

1. Introduction
2. Background
3. Methodology
4. Results
5. Conclusion

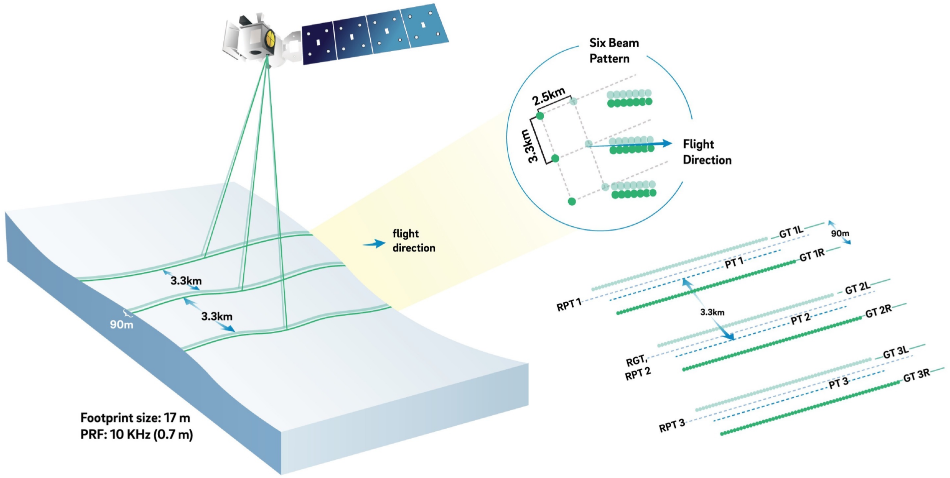


ICESat-2 Satellite (NASA, 2023)

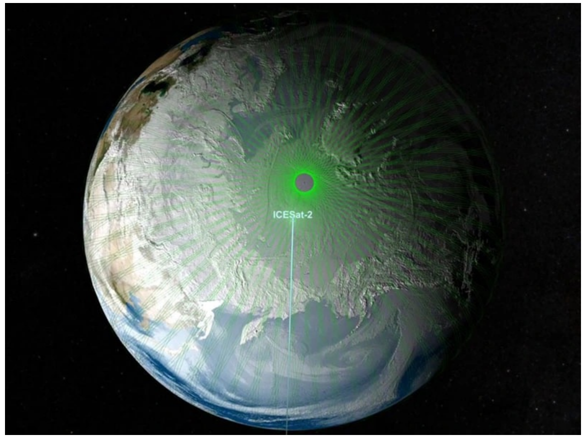
Introduction

The ICESat-2 Satellite Mission

- Data captured from the satellite are sparse
  - 3 pairs of beam (strong and weak beams)
  - 3.3 km between each pair of strong-weak beam
  - 90-100m between each data point along track



ICESat-2 mission beam pattern (Smith et al., 2019)



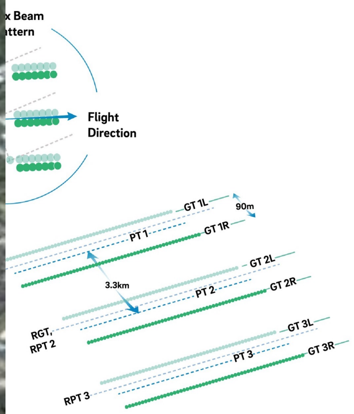
Near-global coverage of ICESat-2 (NASA, 2023)

Introduction

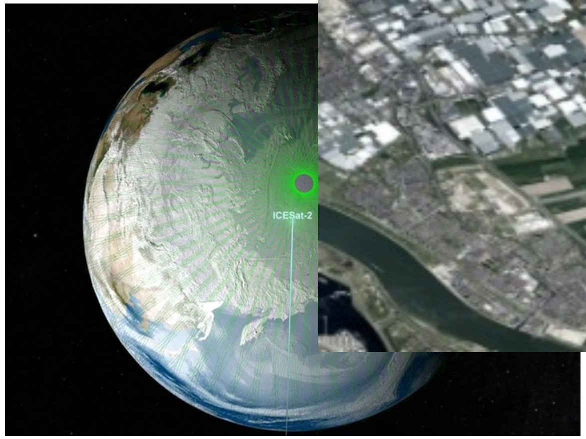
The ICESat-

- Data captu

  - 3 pairs c
  - 3.3 km l
  - 90-100r



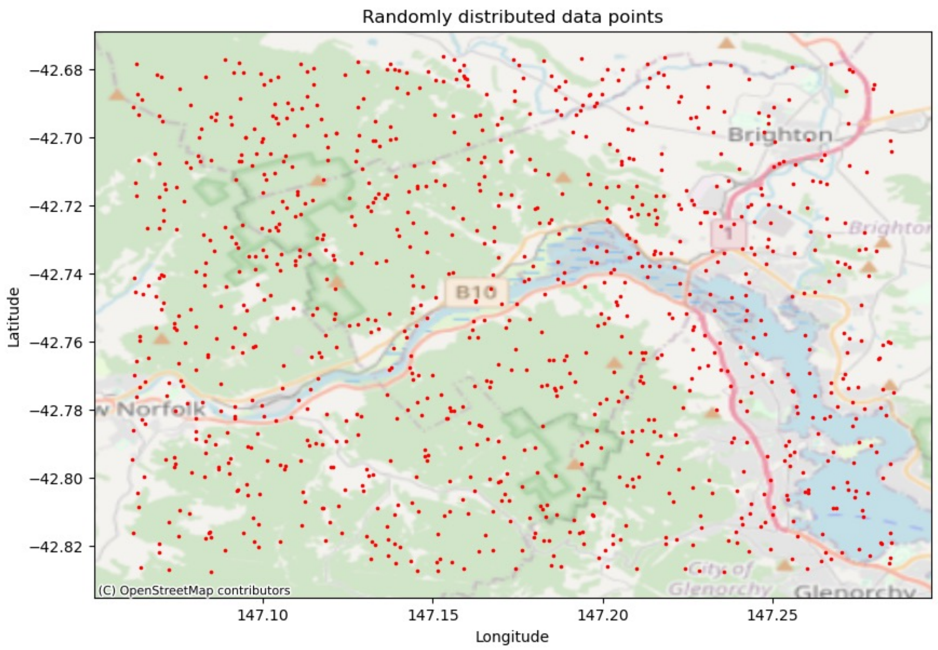
(Smith et al., 2019)



Spatial Interpolation

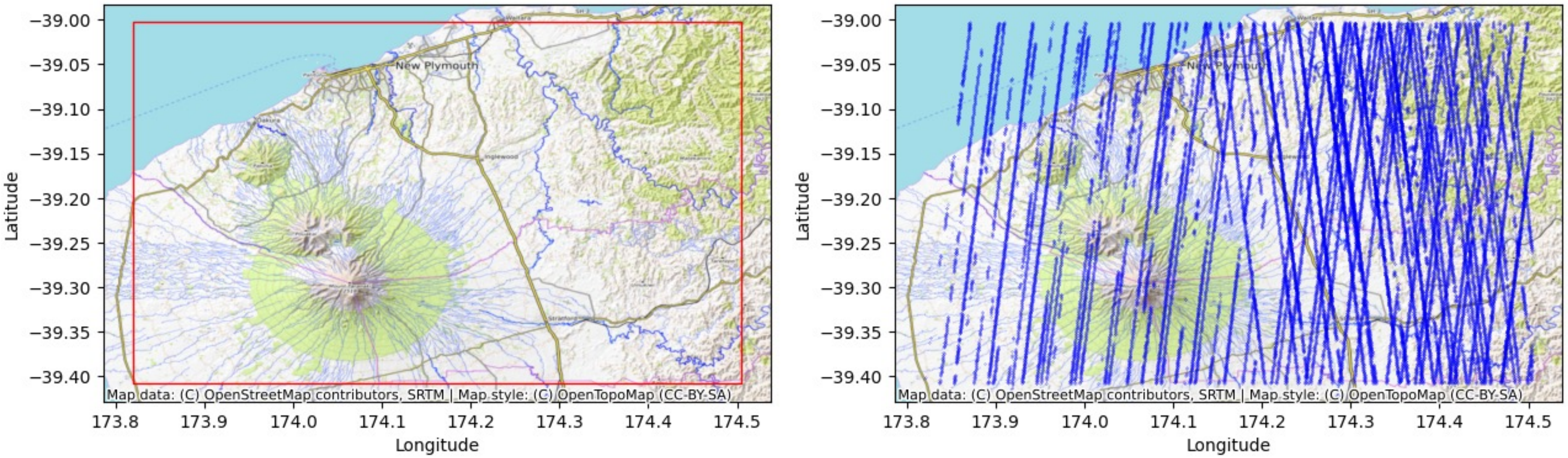
Digital Terrain Modelling

- Estimating values at unsampled locations when there are data gaps
- Used to further perform terrain analysis, run-off modelling, land use planning etc.
- Provides spatial continuity in **2.5D** dimensions (one z value for every x-y coordinate)



Tasmania Samples Dataset (from GEO1015 assignment)

Introduction



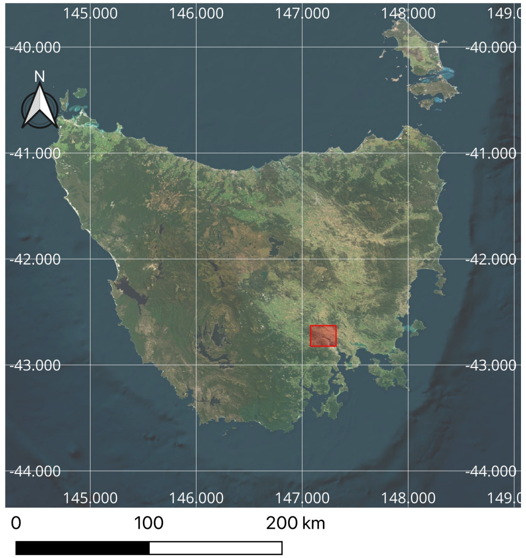
ICESat-2 ATL08 (Land and Vegetation Height) Product with Bounding Box at Mount Taranaki, New Zealand

Research Question

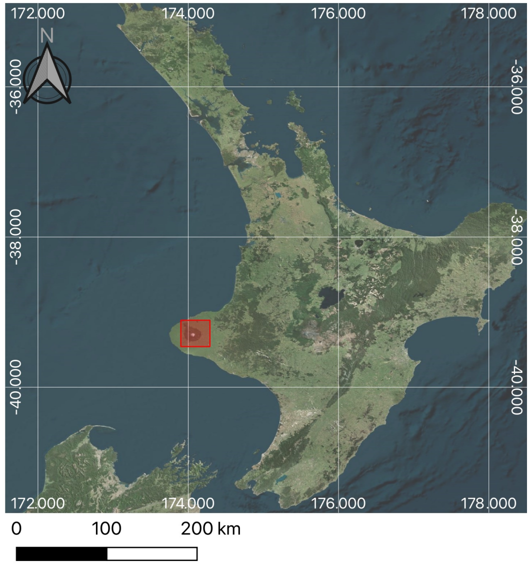
To what extent would Random Forest elevation prediction improve on traditional interpolation methods for creating a DTM?

- Which features in the RF model have the most significant impact on the accuracy?
- How does the accuracy of RF regression vary across different geographical locations within the study areas?
- How would the RF model that created in this research compared against the method proposed in RFsp by Hengl et al. [2018]?

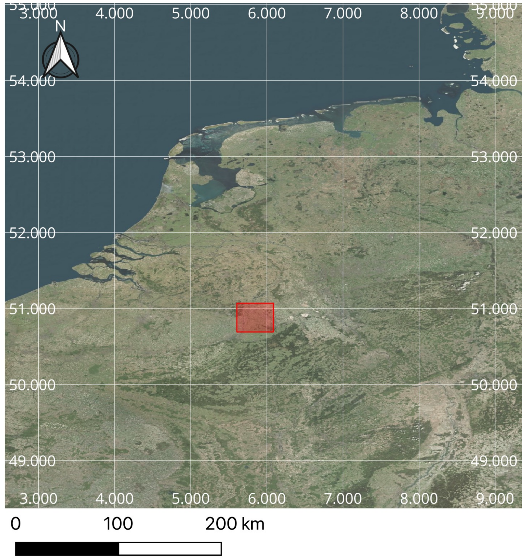
Study Areas



Tasmania
Australia



Mount Taranaki
New Zealand



Limburg
Netherlands

Grand Canyon
United States

Background

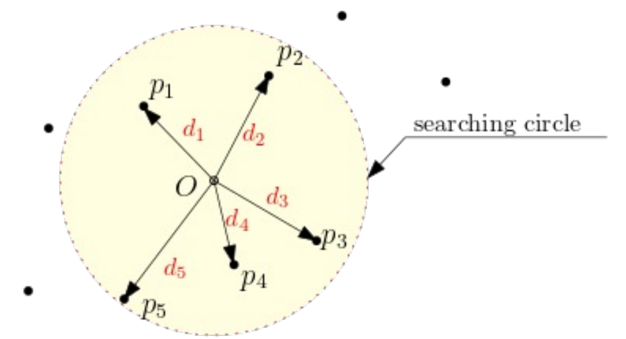
Traditional Interpolation

- Inverse Distance Weighting (IDW)
- Triangular Irregular Network (TIN)
- Laplace Interpolation
- Natural Neighbour Interpolation

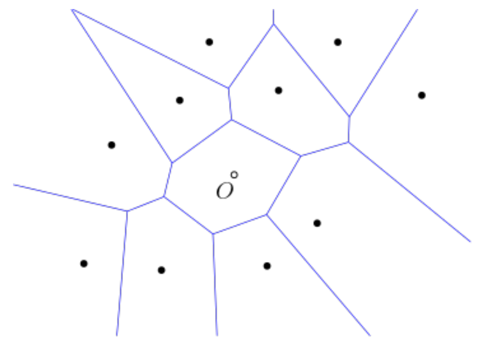
Random Forest (RFsp)

- Random Forest as a Generic Framework (Hengl et. al., 2018)
- 'Covariates' == Features

$$Y(s) = f(X_G, X_R, X_P)$$

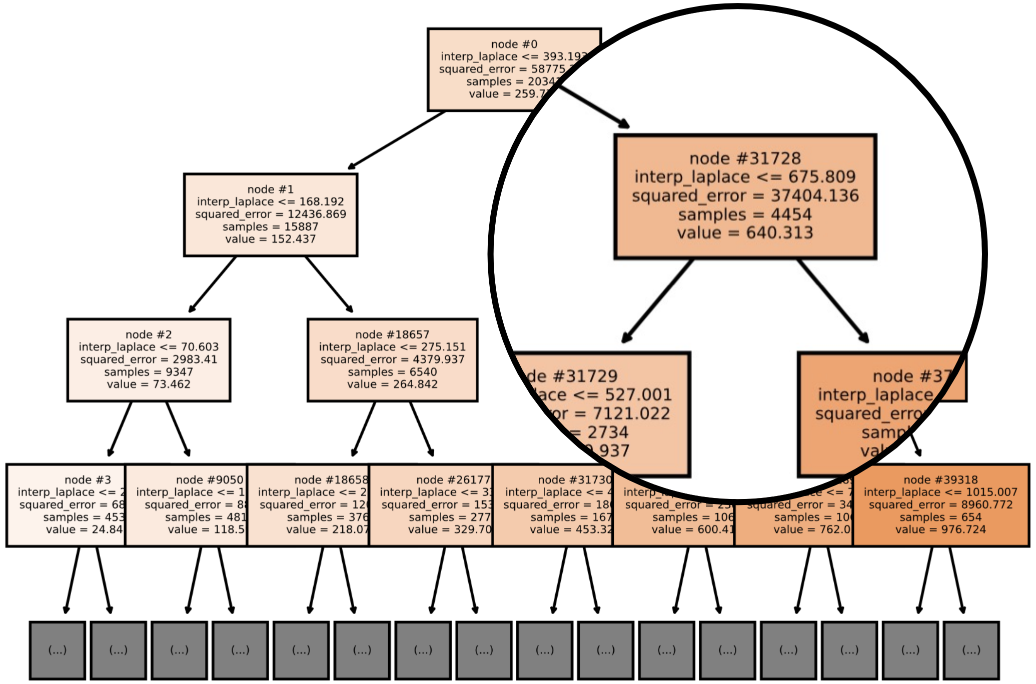


IDW interpolation



Voronoi Diagram

Random Forest Regression

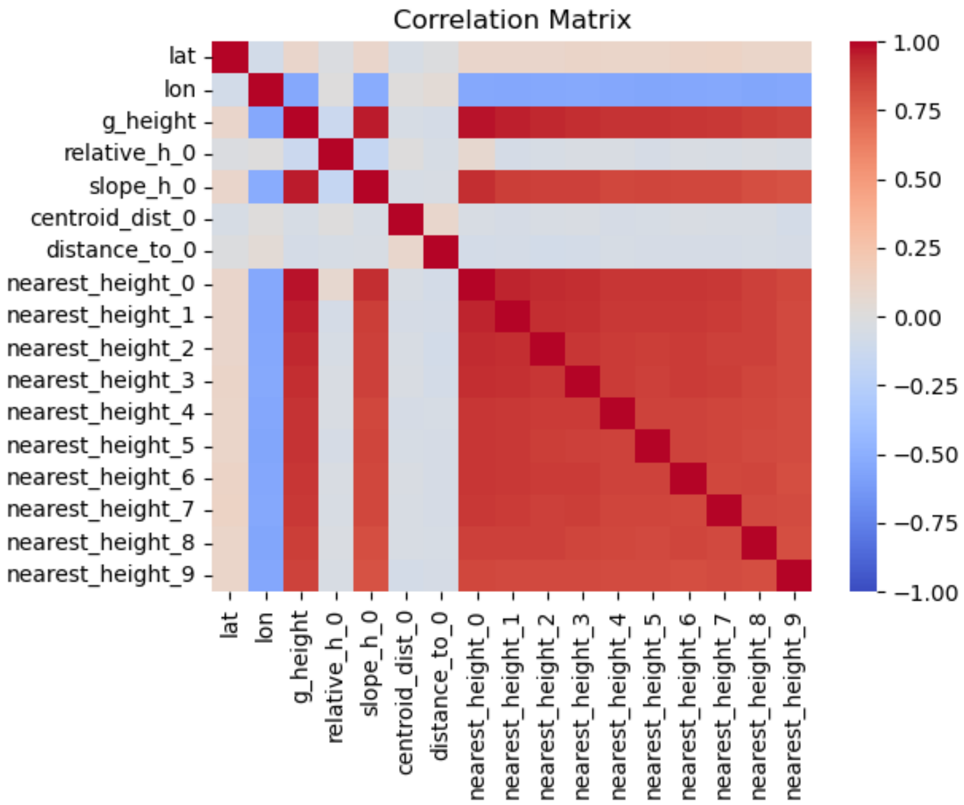


- Random forest will use a lot of decision trees to decide the height of the terrain using aggregation and bagging
- Using auxiliary data known as ‘**Features**’ to train the random forest
- Each node split automatically left or right resulting lower loss function based on a random subset of training data

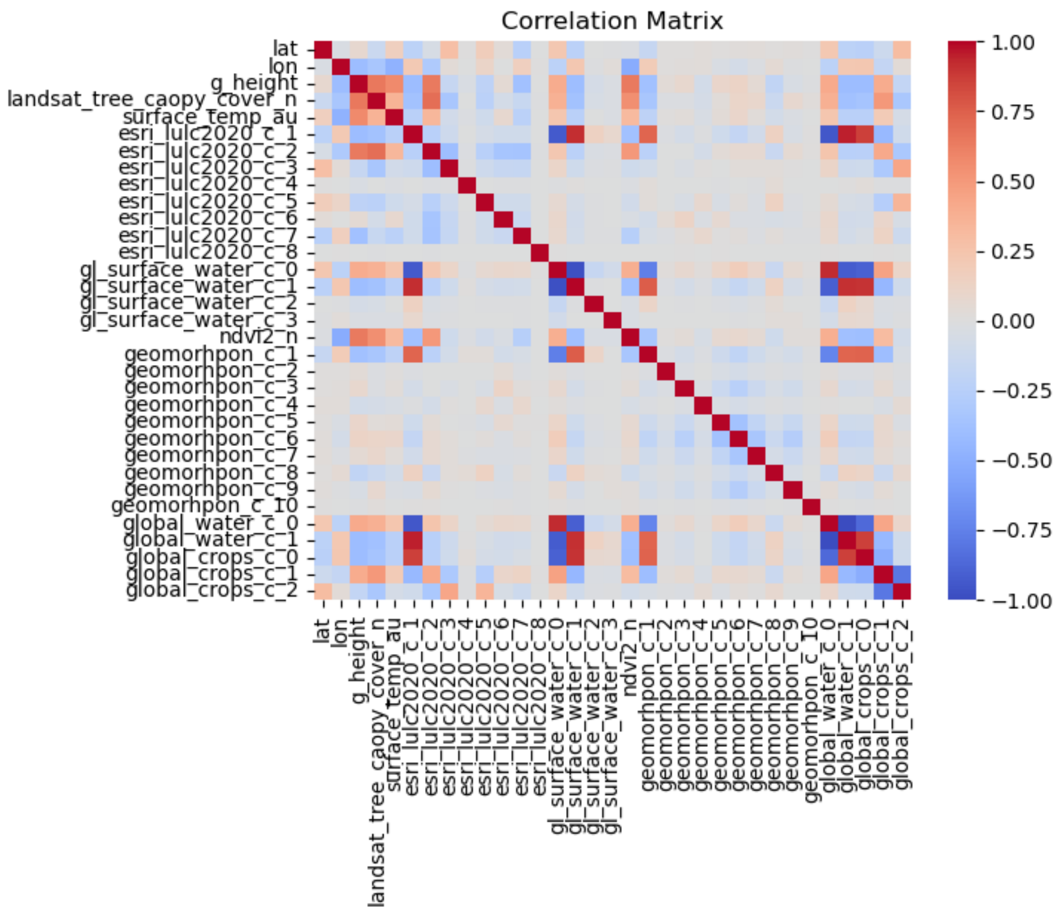
Features Used in this study

1. **Geometric Features**
   - Distance to the nearest ICESat-2 Points
   - Relative height with respect to ICESat-2 n-number of points
   - Slope with respect to ICESat-2 n-number of points
   - Neighbour height of nearest ICESat-2 point
2. **Remote Sensing Features**
   - Land Use/Land Cover
   - Water Bodies Mask
   - World Settlement Footprint
   - Normalized difference vegetation index (NDVI)
   - Cropland Classification

Correlation Map



(1) Geometric Features



(2) Remote Sensing Features

Accuracy Assessment

- **Mean Absolute Error (MAE)**

  MAE quantifies the average absolute difference between predicted and ground truth values.

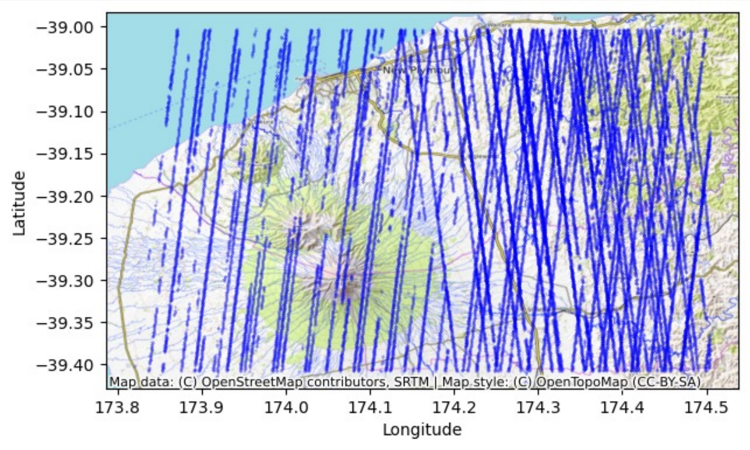
- **Root Mean Square Error (RMSE)**

  RMSE quantifies the average difference between predicted and ground truth values by taking the square root of the mean of squared differences.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - z_i)^2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{z}_i - z_i\right|$$

Methodology



ICESat-2 Data:
Latitude,
Longitude,
Ground Truth Height

Features: One-hot encoding

Categorical data to Binary (1 and 0)

Each category belongs to one column

Easier for model to process categorical information effectively

id	LULC
0	Trees
1	Built Area
2	Built Area
3	Crops
4	Built Area
5	Trees

id	LULC_Trees	LULC_Built_Area	LULC_Crops
0	1	0	0
1	0	1	0
2	0	1	0
3	0	0	1
4	0	1	0
5	1	0	0

Traning of the Model

Lat	Lon	Height	Features

Target Varaible

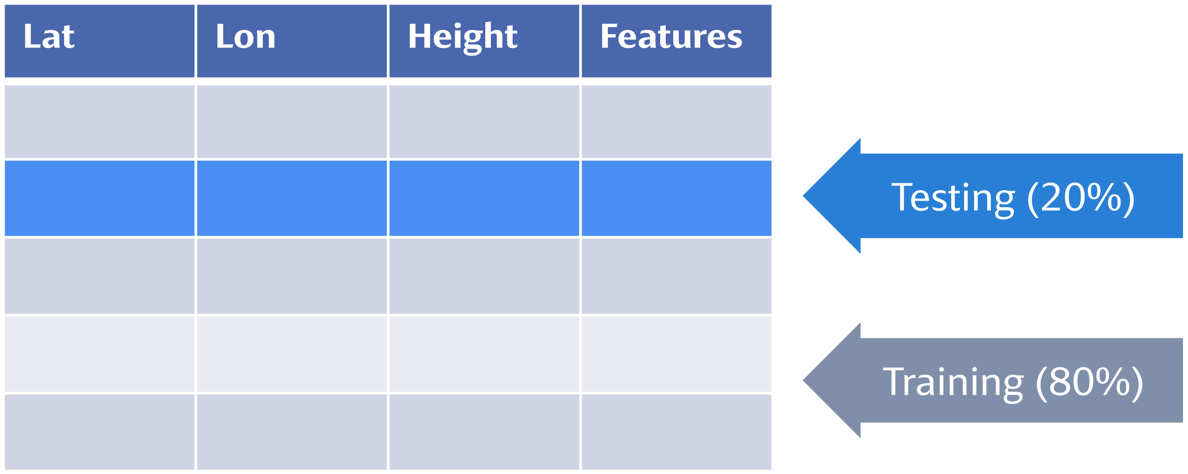
Geometric Features

- Distance to ICESat-2
- Nearest Neighbour Height
- Gradient
- Relative Height

Remote Sensing Features

- Water Mask
- World Settlement
- Cropland
- Geomorphons
- Land Use/Land Cover
- NDVI
- Earth Surface Spectural
- Forest Coverage

Test-Train Split



Test-Train Split

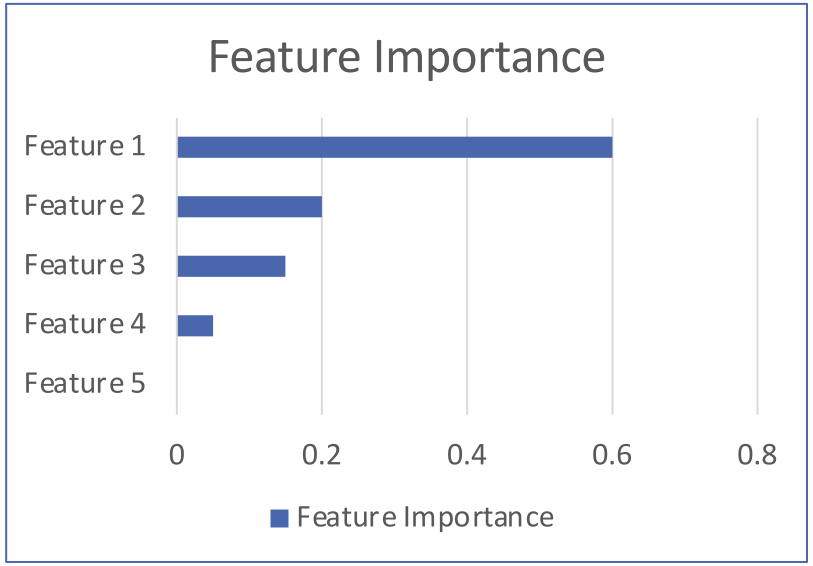
Evaluation on how well the model performs on unseen data.
Randomly selected from the data frame

- ⇨ 80% Testing Data
- ⇨ 20% Training Data

Feature Importance

Mean Decrease of Accuracy

- Using Means Squared Error as accuracy metric
- Each values are shuffled randomly
- It evaluates how each feature contributed to the RF model by decrease of accuracy



Results

Test Geometric Features

- Test 1: Distance to Nearest ICESat-2 Point
- Test 2: Nearest Neighbour Height
- Test 3: Gradient to neighbours
- Test 4: Relative Height

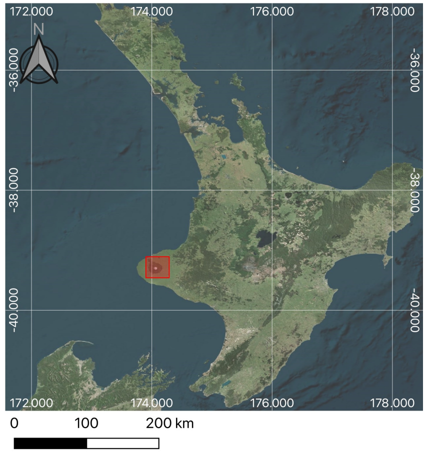
Test Remote Sensing Features

- Test 5: Remote Sensing Features

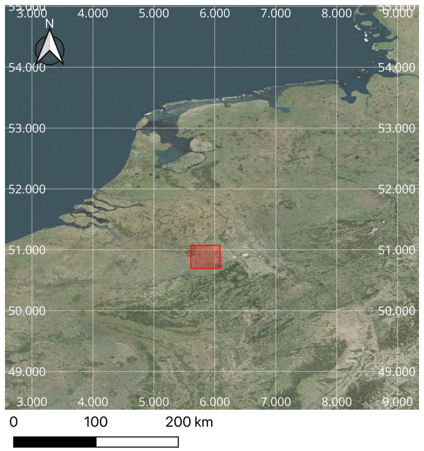
Test Combined Models from other Study Areas

- Combine models from Mount Taranaki, Grand Canyon and Limburg
- Apply new model on Tasmania, Australia

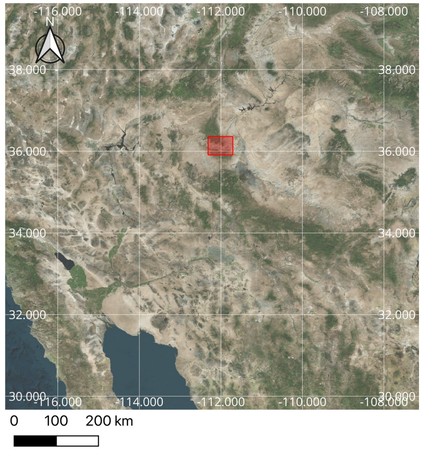
Combinding RF Models



Mount Taranaki
New Zealand



Limburg
Netherlands

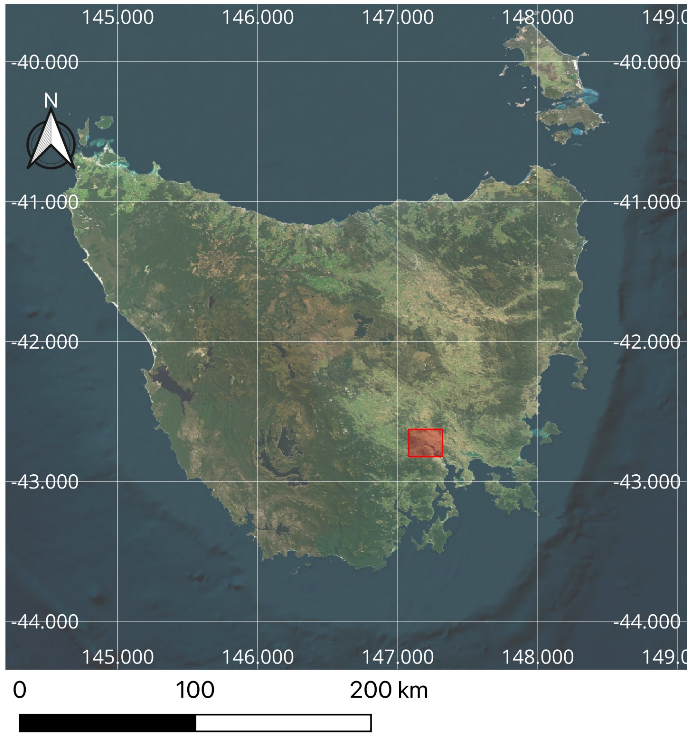


Grand Canyon
United States

Model

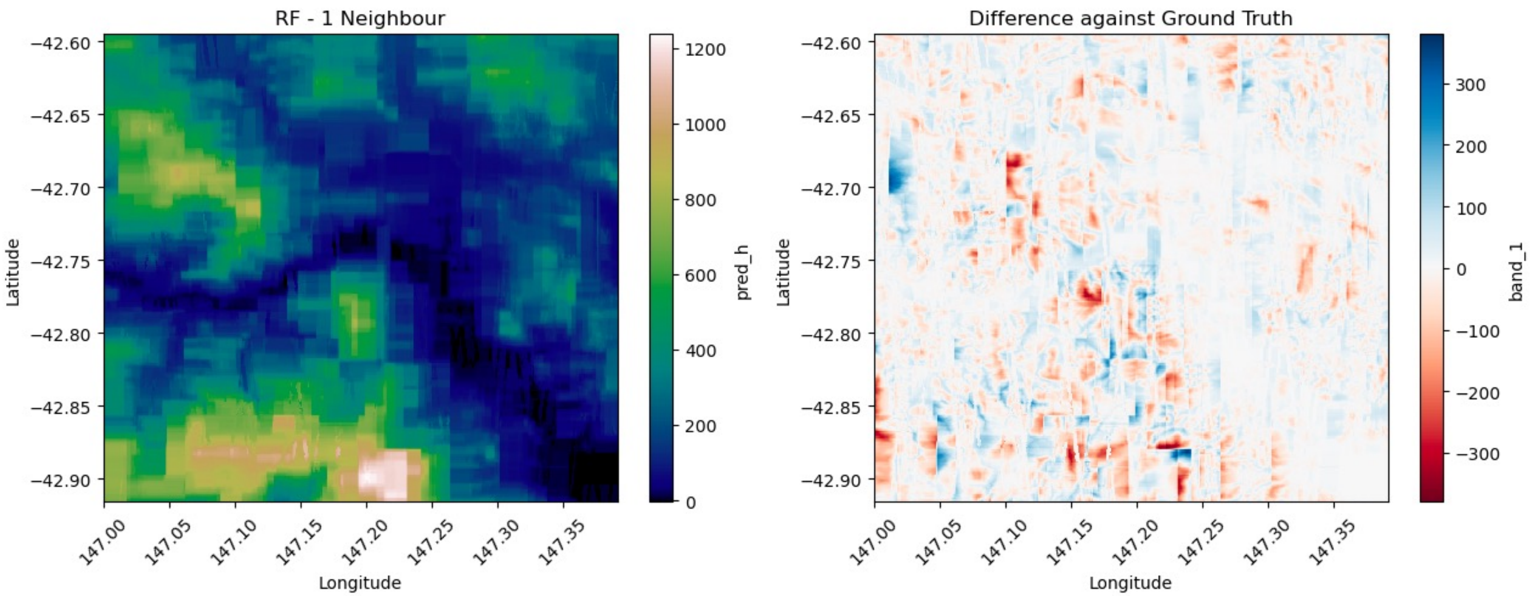
Model

Model



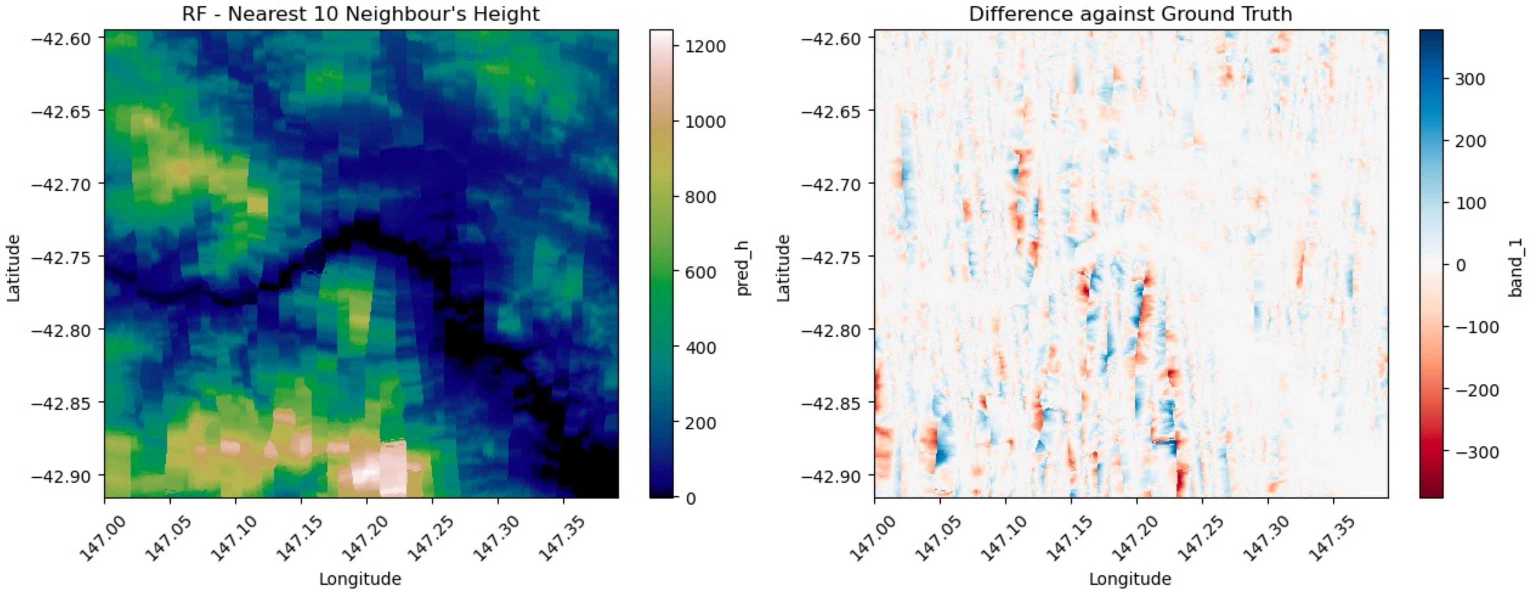
Tasmania
Australia

Test 1: Distance to Nearest ICESat-2 Point



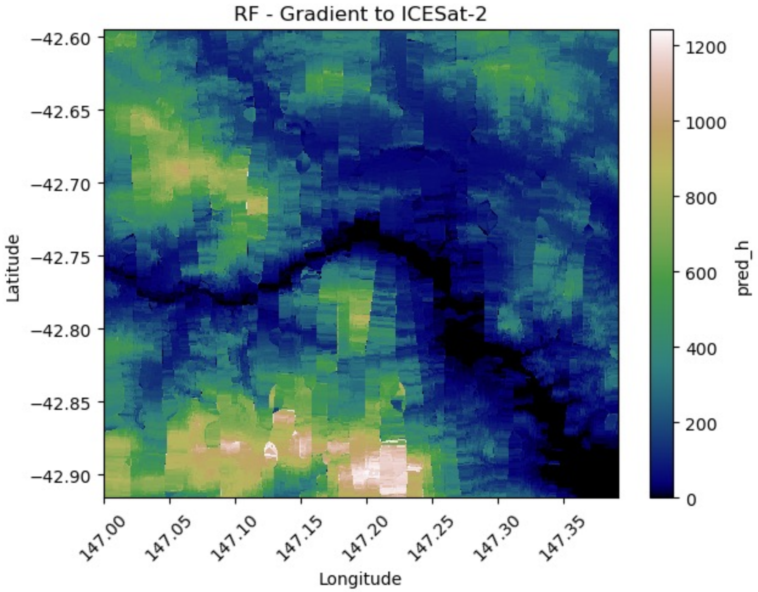
N-neighbours	Min Diff	Max Diff	RMSE	MAE
1	-379.611	375.658	49.916	31.051
2	-372.352	352.439	50.075	31.276
5	-390.528	424.799	54.068	34.91
10	-398.333	371.783	56.325	36.784
100	-395.529	571.263	68.555	44.677

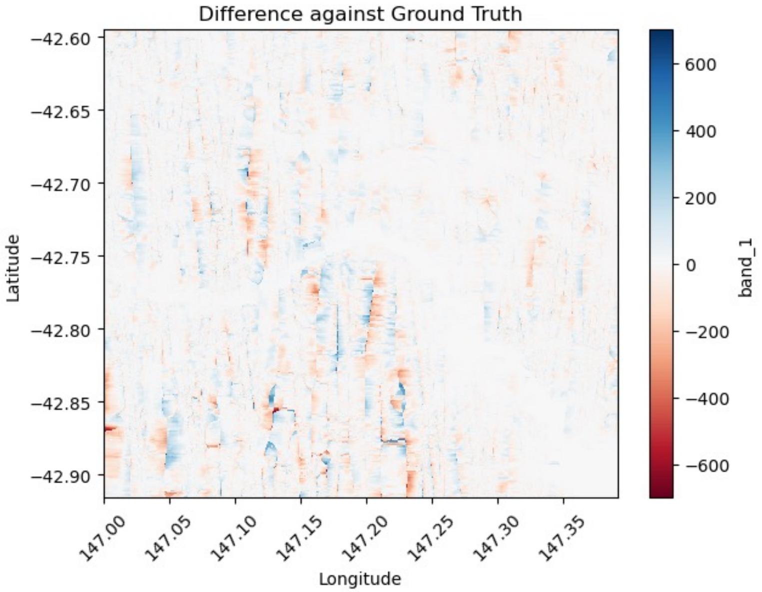
Test 2: Nearest Neighbour Height



N-neighbours	Min Diff	Max Diff	RMSE	MAE
1	-344.246	385.154	45.688	27.079
2	-341.152	387.564	44.025	25.242
10	-341.805	376.818	43.63	24.923
100	-337.345	377.037	43.737	25.06
200	-335.816	374.136	43.753	25.078

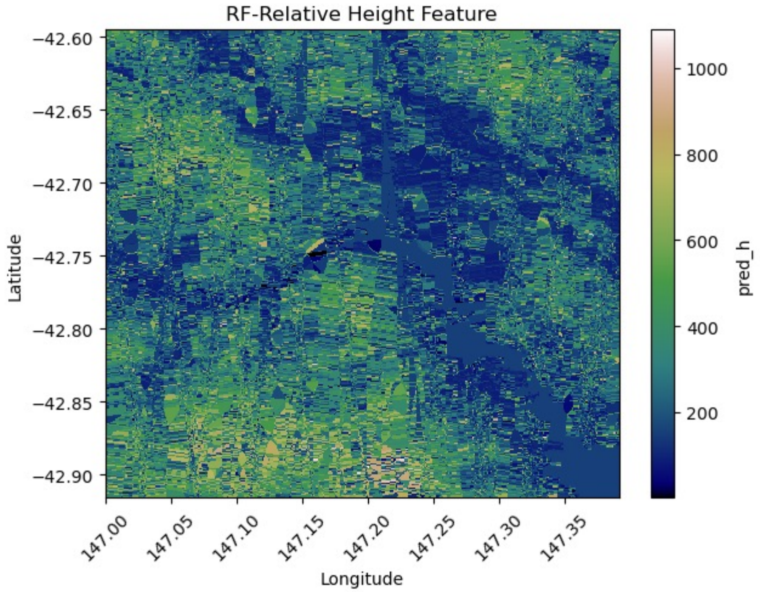
Test 3: Nearest Gradient

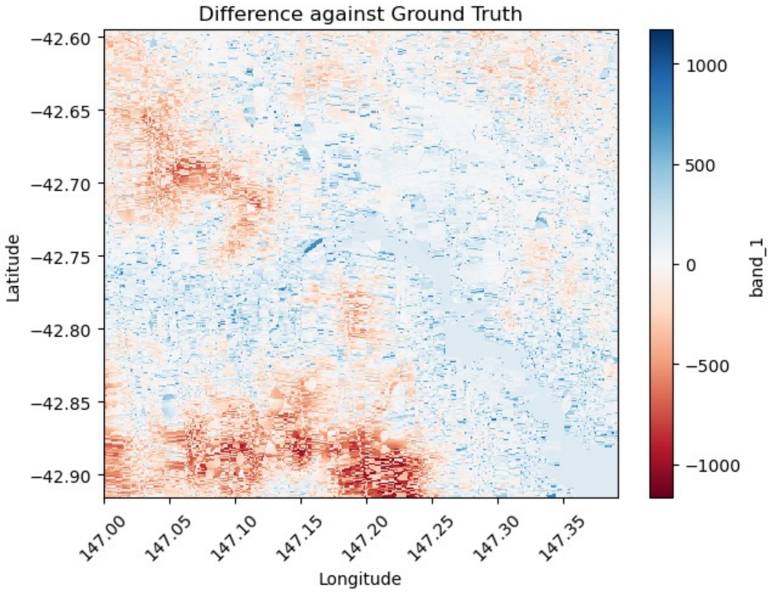




N-neighbours	Min Diff	Max Diff	RMSE	MAE
1	-391.77	394.306	47.577	29.076
5	-680.321	663.468	54.827	31.094
10	-594.567	627.791	55.565	31.742

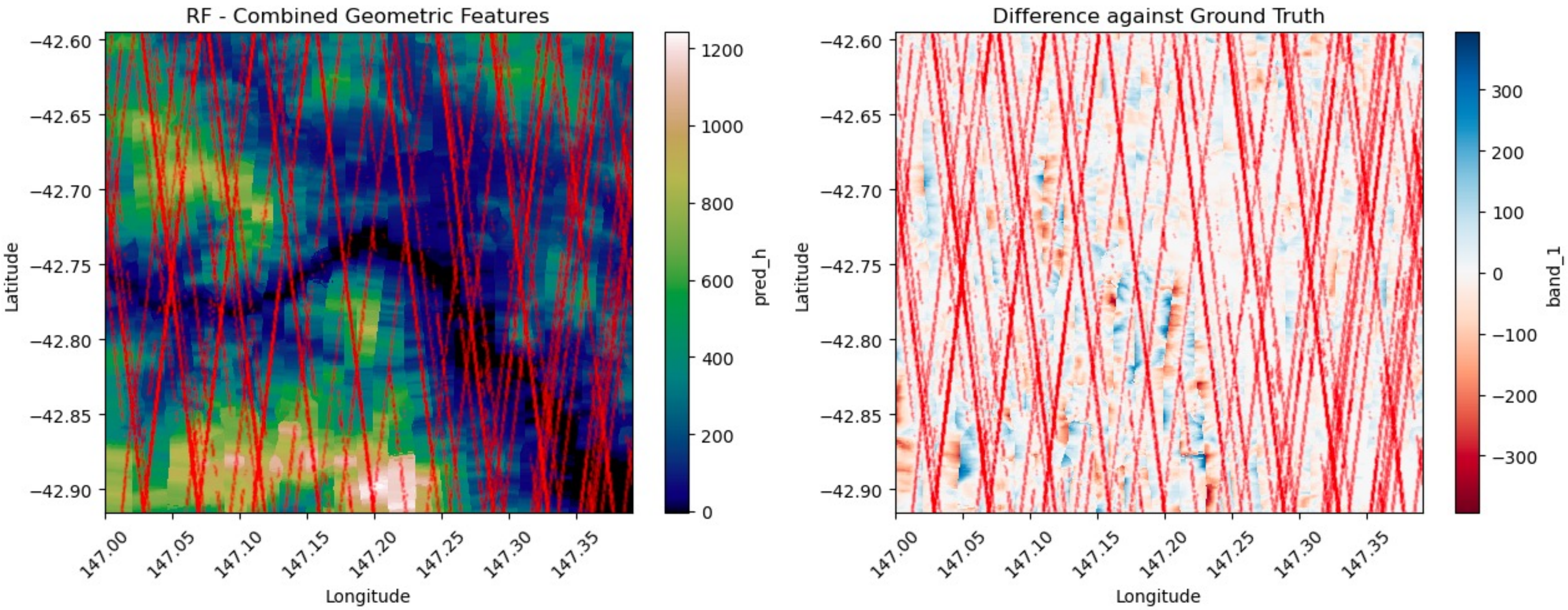
Test 4: Relative Height





N-neighbours	Min Diff	Max Diff	RMSE	MAE
1	-353.554	340.476	47.573	29.034
2	-390.509	334.424	48.724	30.545
10	-787.84	564.375	58.619	37.006
50	-773.813	550.006	69.181	42.806
100	-885.601	494.715	76.984	45.242

Combined Geometric Features



Tests on RF Features

Test Geometric Features

- Test 1: Distance to Nearest ICESat-2
- Test 2: Nearest Neighbour Height
- Test 3: Slope to neighbours
- Test 4: Relative Height

Test Remote Sensing Features

- Test 5: Remote Sensing Features

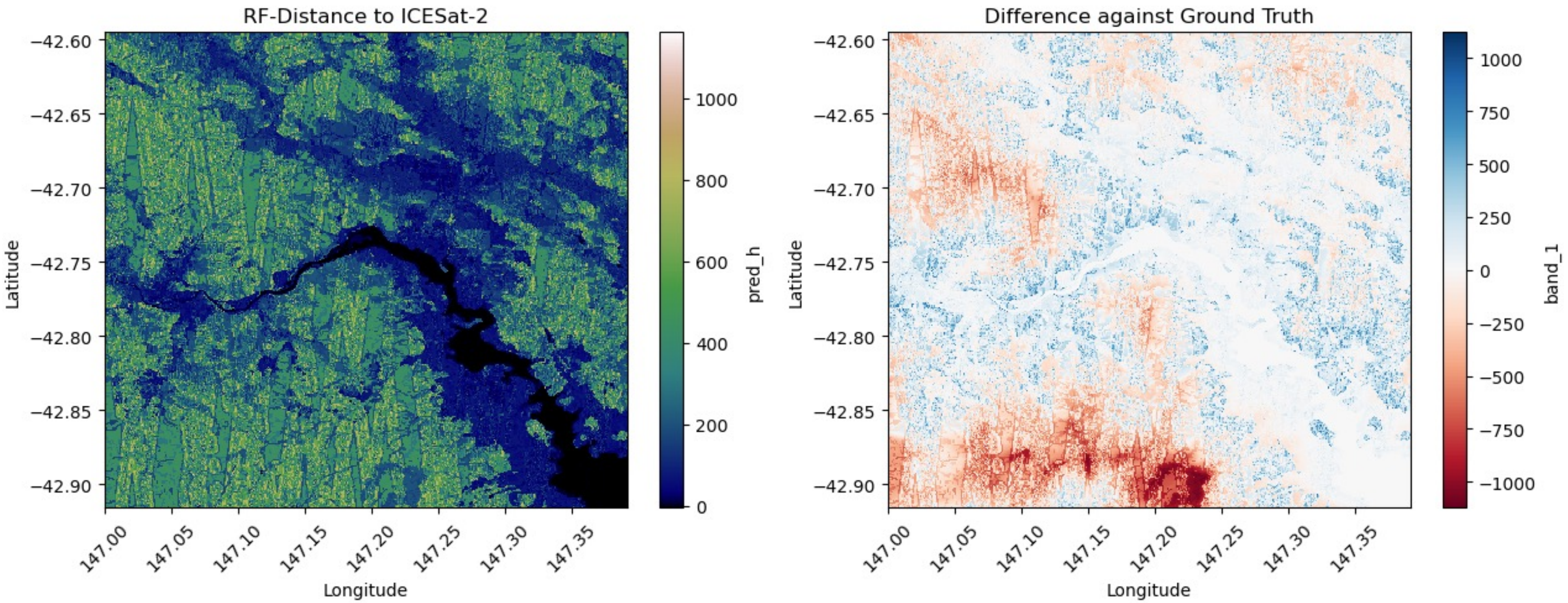
1 nearest-neighbour

10 nearest-neighbour

5 nearest-neighbours

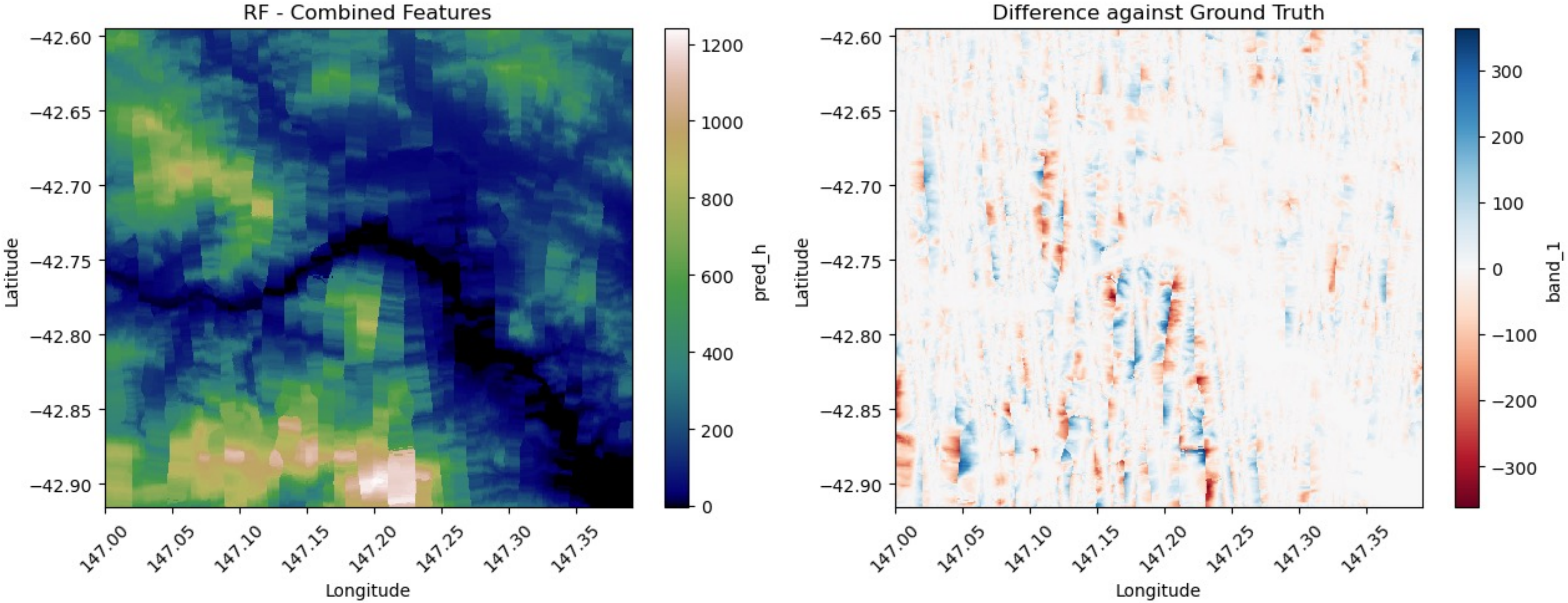
1 nearest-neighbour

Test 5: Remote Sensing Features

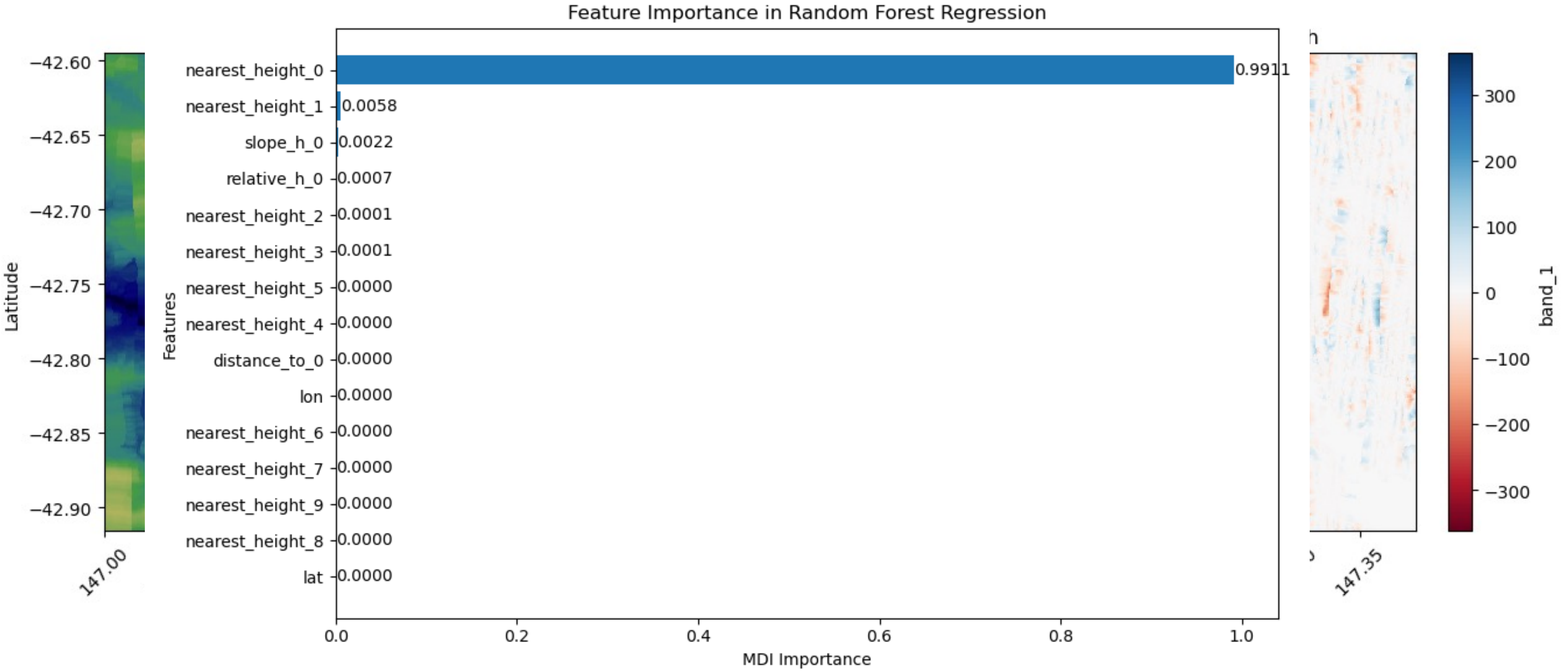


Nearest Distance | Water Mask | Geomorphon | NDVI Features

Tasmania: Combing Features

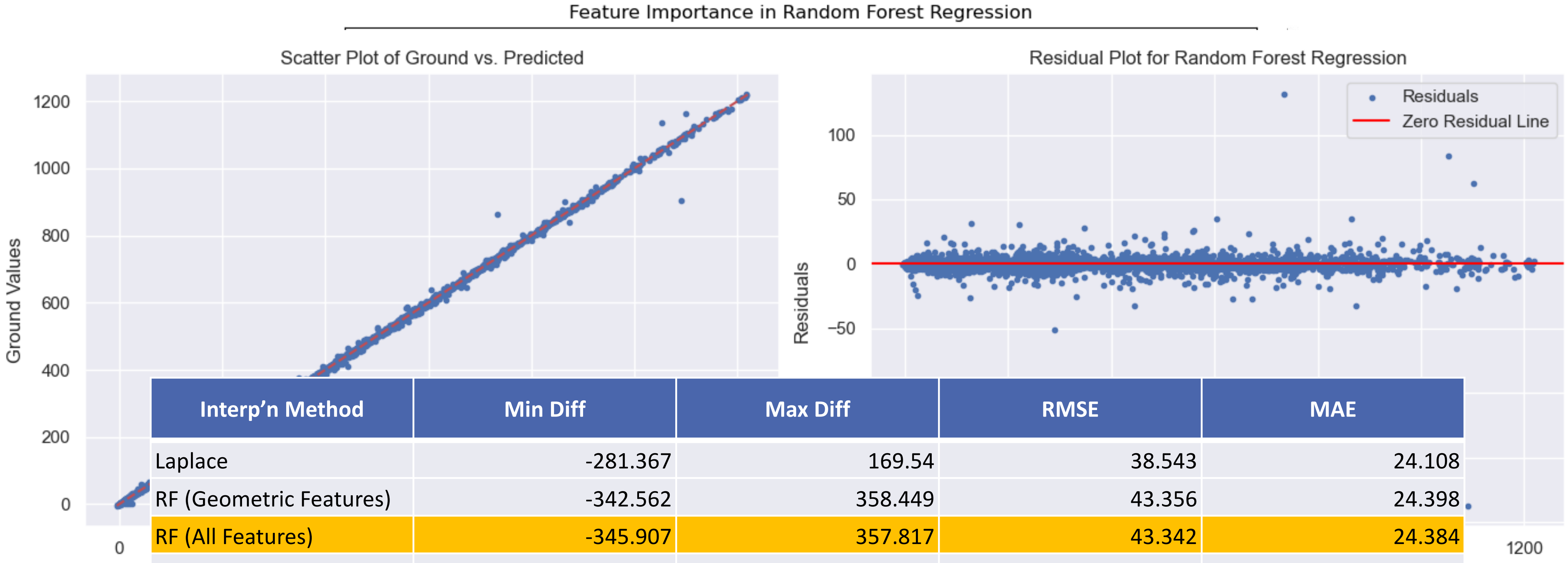


Tasmania: Combing Features



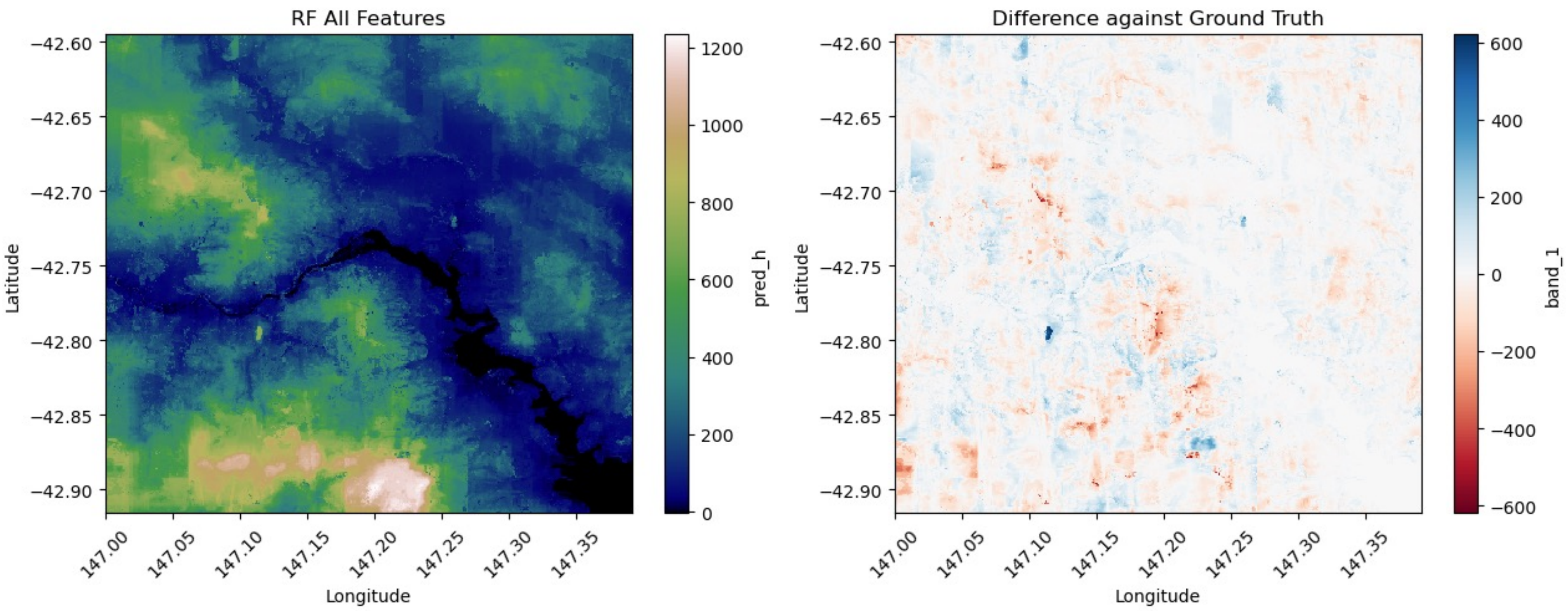
Tasmania: Combing Features

Tasmania: Combing Features

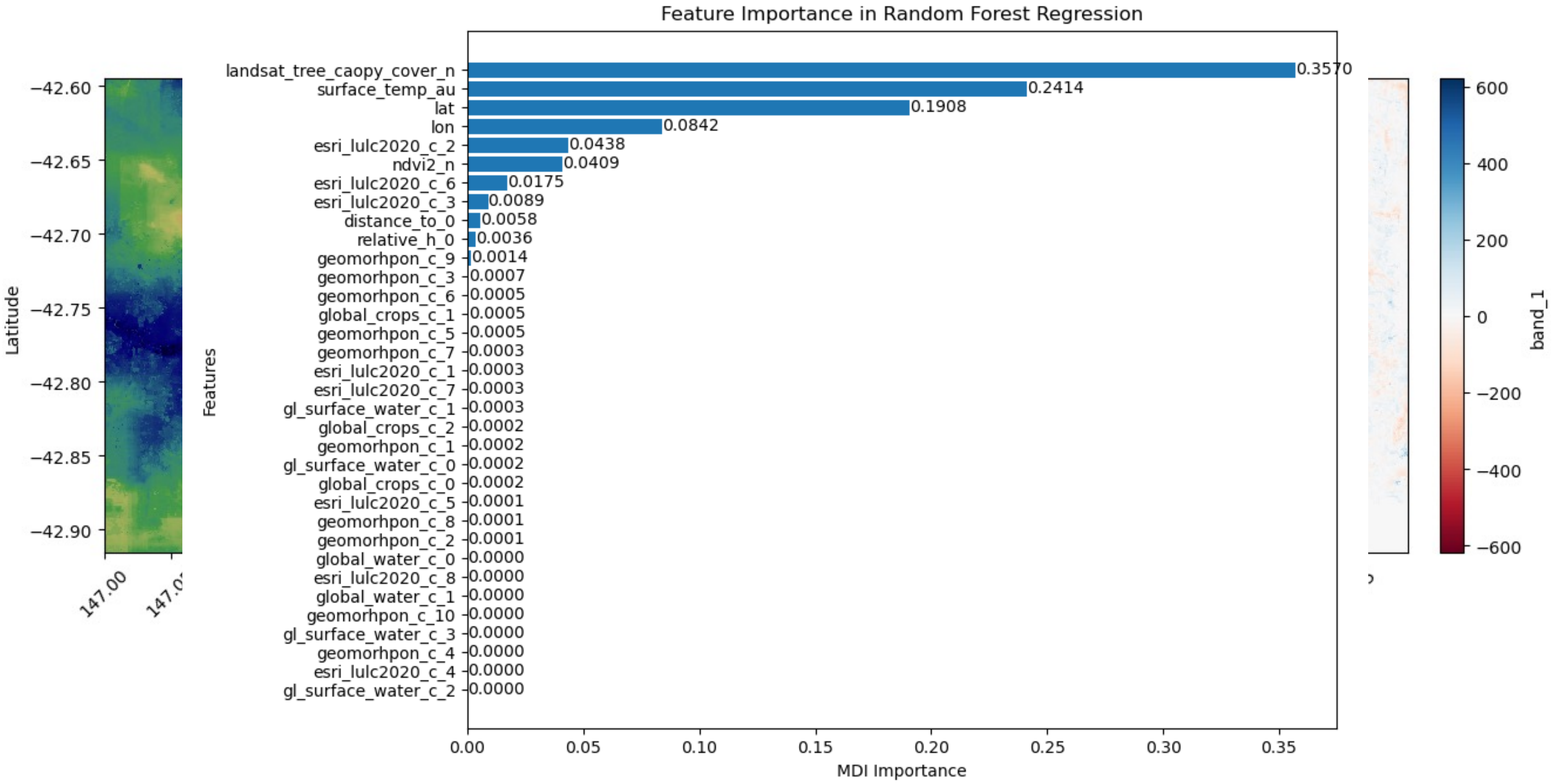


Interp'n Method	Min Diff	Max Diff	RMSE	MAE
Laplace	-281.367	169.54	38.543	24.108
RF (Geometric Features)	-342.562	358.449	43.356	24.398
RF (All Features)	-345.907	357.817	43.342	24.384
RF (All Features, excl. Nearest Neighbour Height)	-597.919	620.033	55.711	34.986

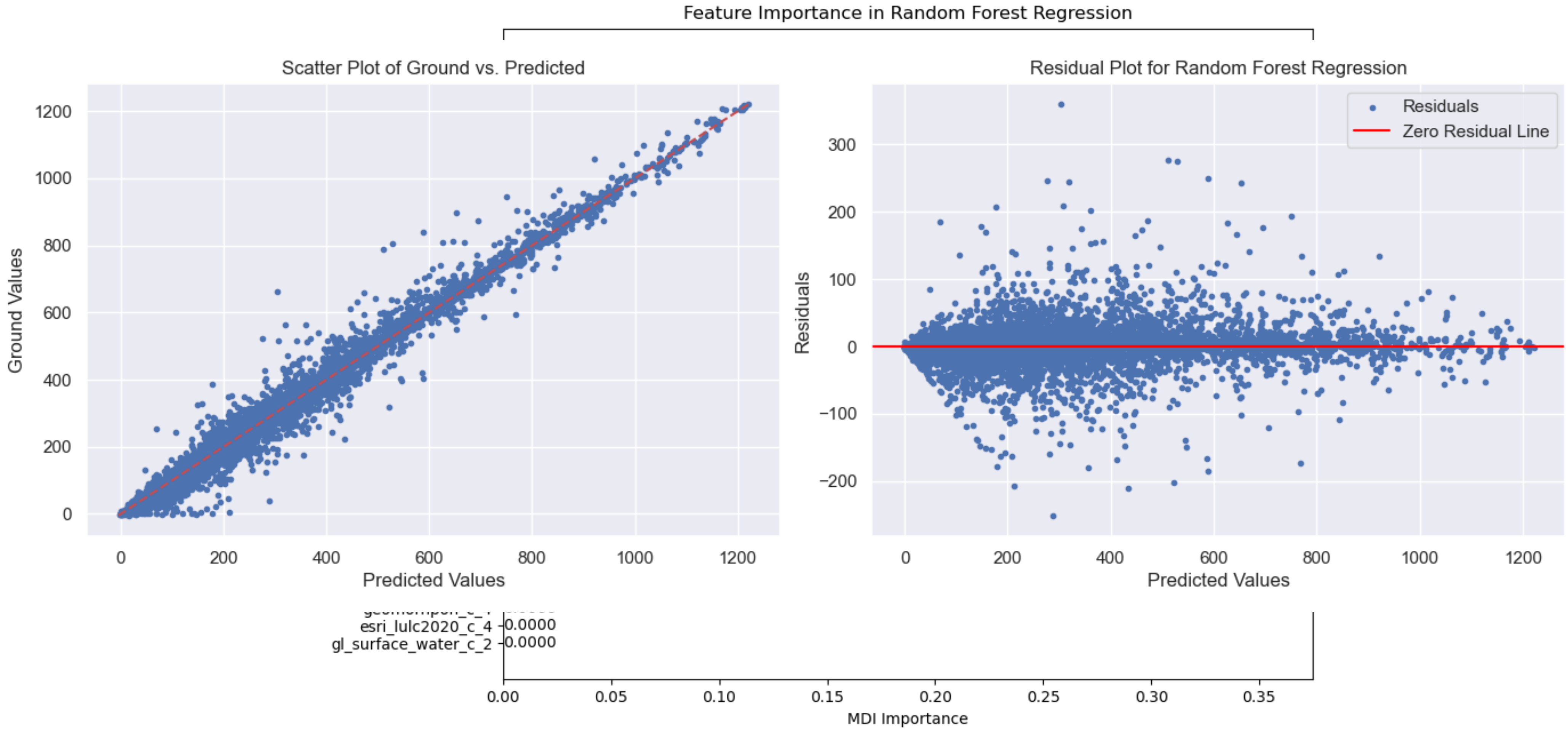
Tasmania: Combing Features



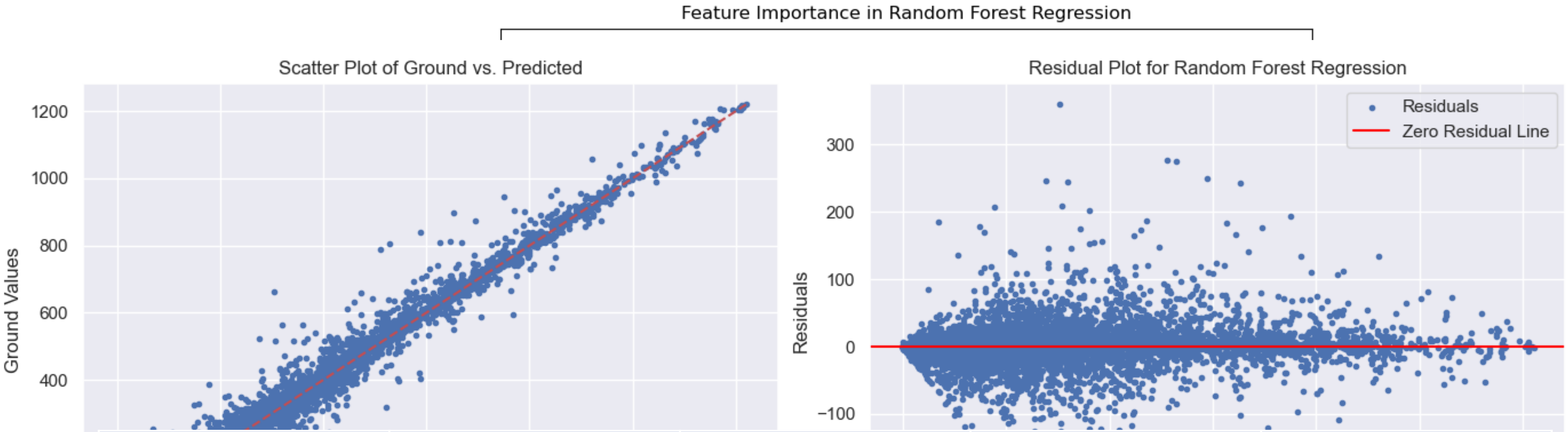
Tasmania: Combing Features



Tasmania: Combing Features



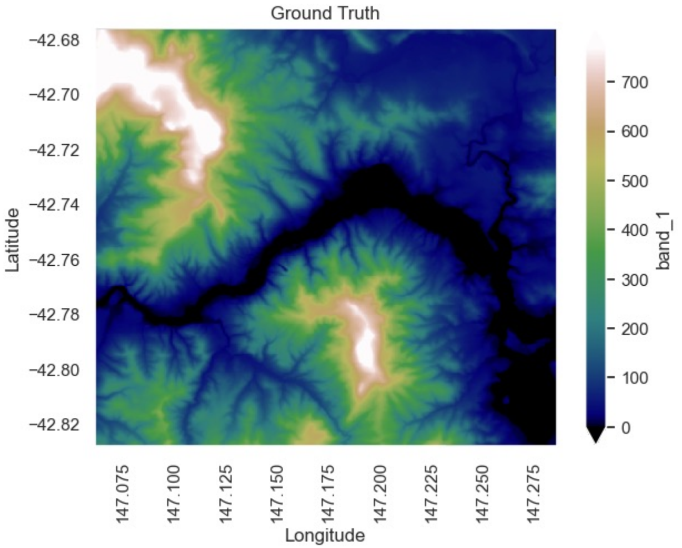
Tasmania: Combing Features

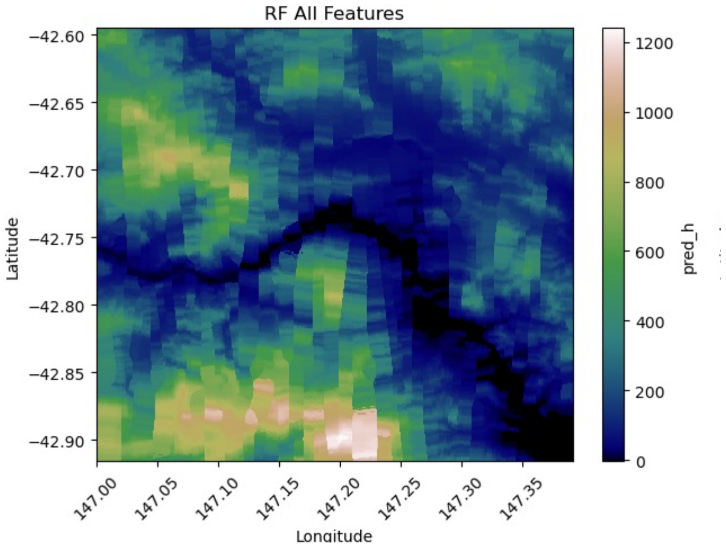


Interp'n Method	Min Diff	Max Diff	RMSE	MAE
Laplace	-281.367	169.54	38.543	24.108
RF (Geometric Features)	-342.562	358.449	43.356	24.398
RF (All Features)	-345.907	357.817	43.342	24.384
RF (All Features, excl. Nearest Neighbour Height)	-597.919	620.033	55.711	34.986

Comparing DTMs

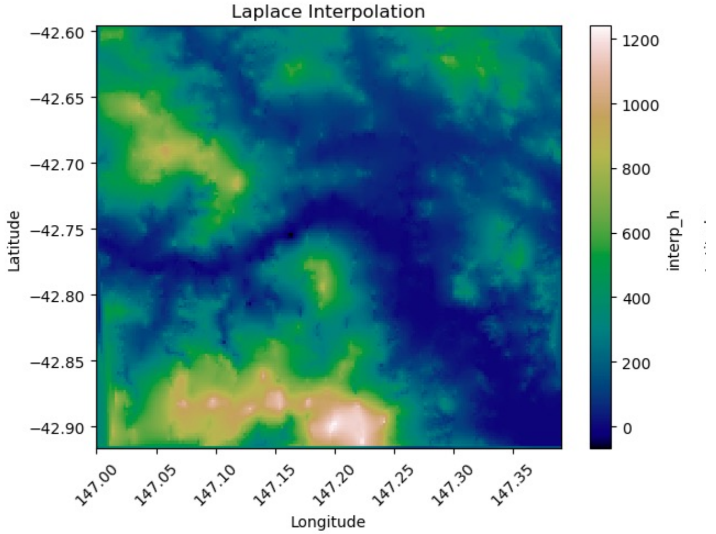
Ground Truth

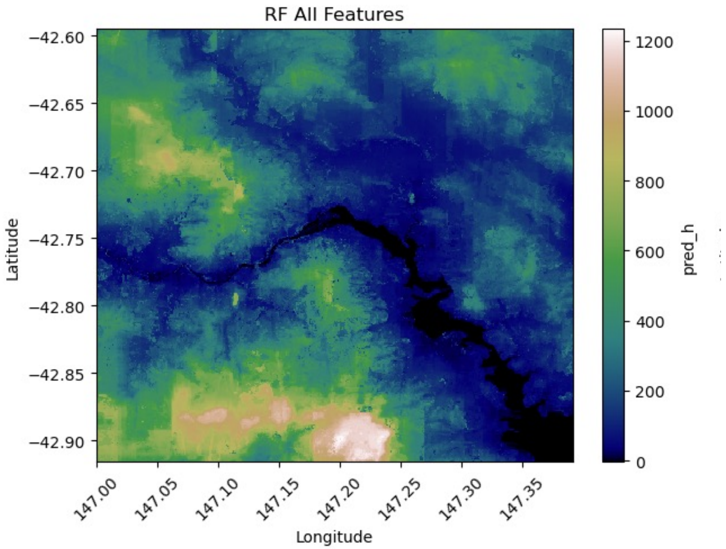




RF (All Features)

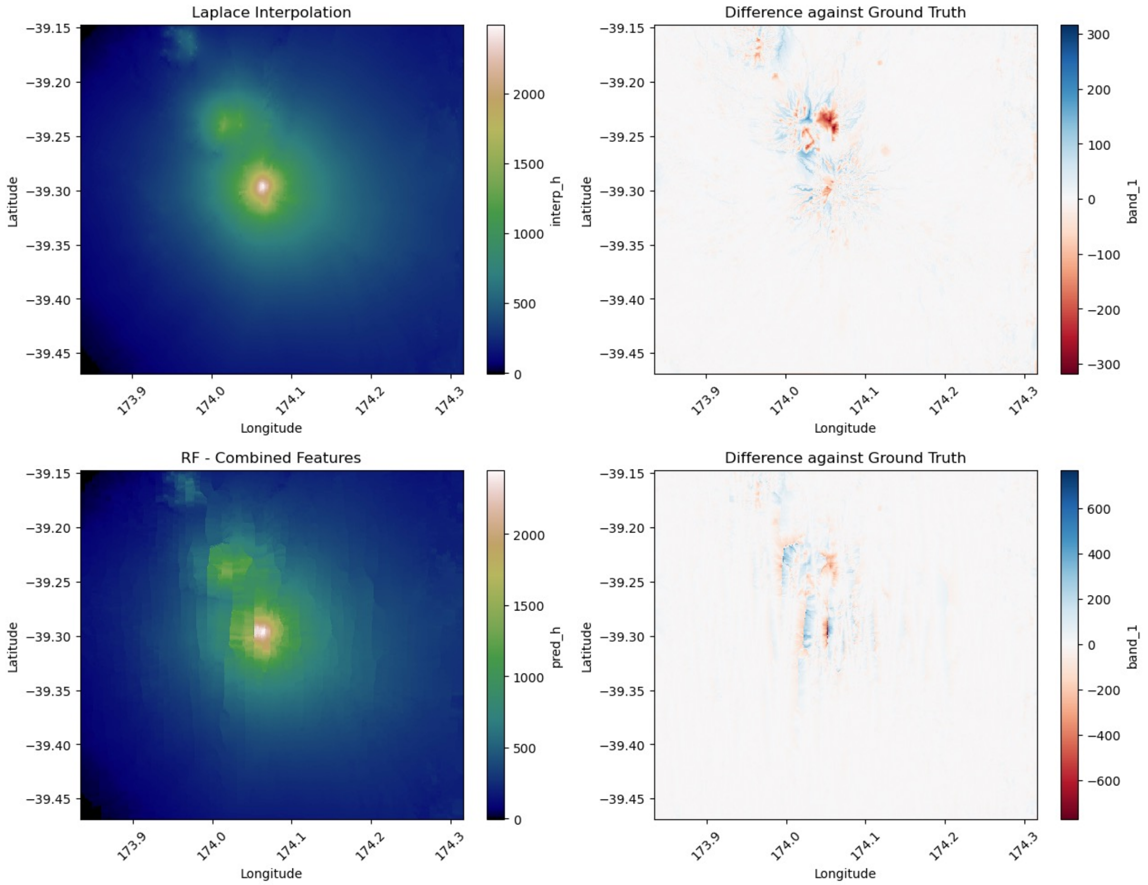
Laplace Interp’n





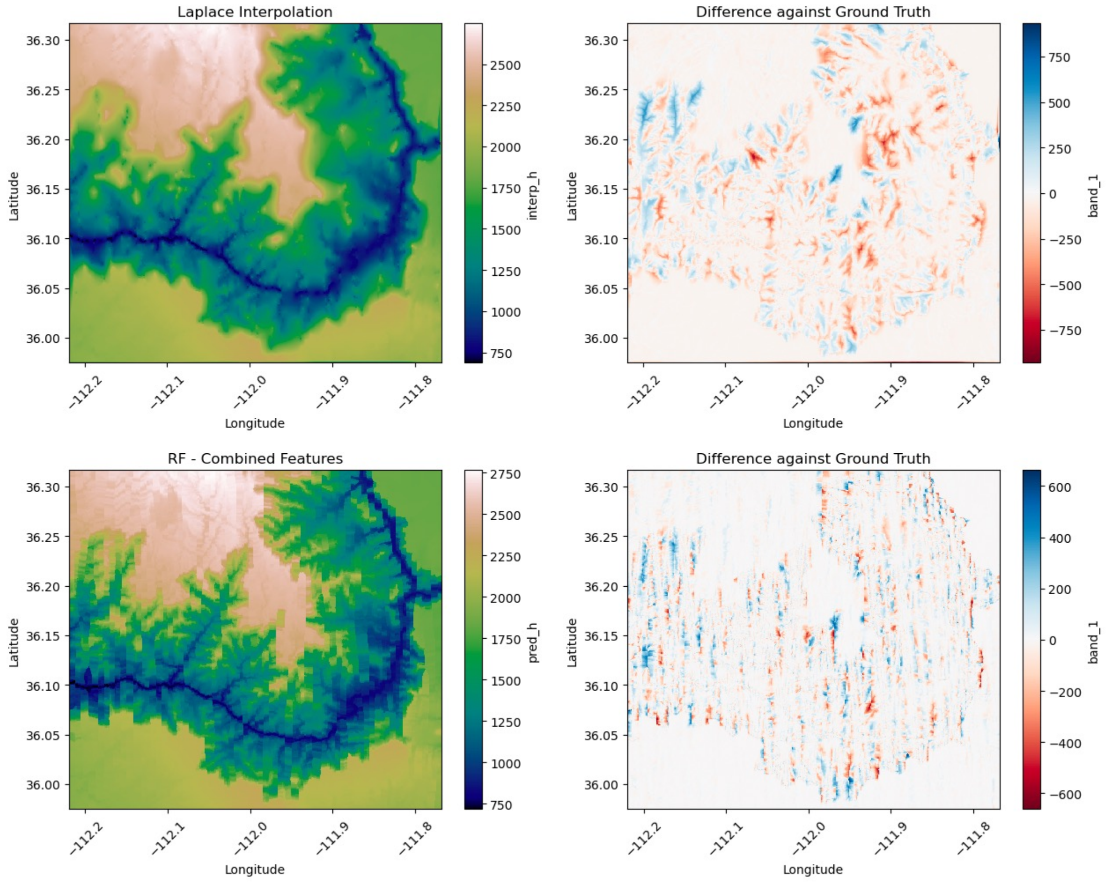
RF (All Features excl. Nearest Neighbour Height)

Geographical Locations



Mt Taranaki	Min Diff	Max Diff	RMSE	MAE
Laplace	-317.31	216.004	16.11	6.451
RF (All Features)	-770.826	526.337	28.424	11.092

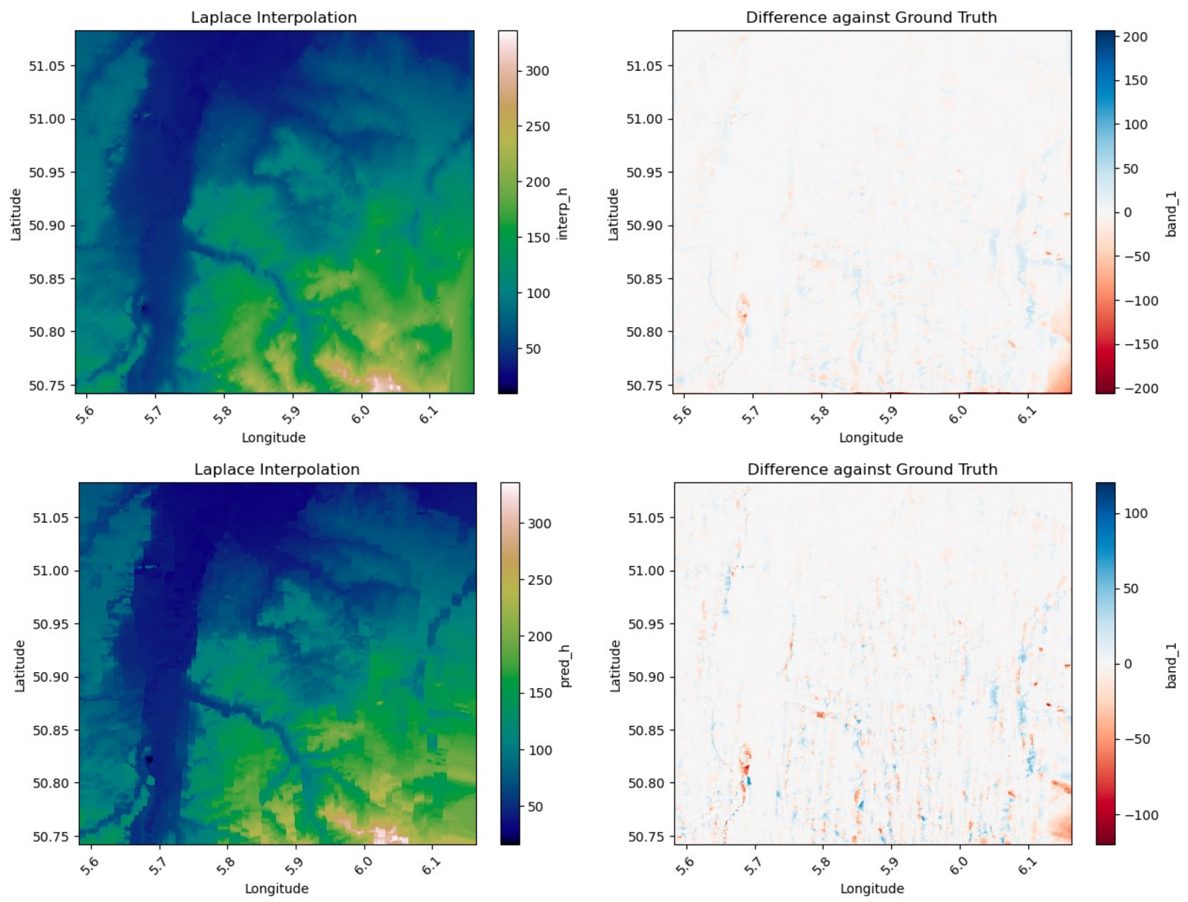
Geographical Locations



Mt Taranaki	Min Diff	Max Diff	RMSE	MAE
Laplace	-317.31	216.004	16.11	6.451
RF (All Features)	-770.826	526.337	28.424	11.092

Grand Canyon	Min Diff	Max Diff	RMSE	MAE
Laplace	-792.423	931.947	109.017	69.526
RF (All Features)	-653.547	661.584	76.177	39.448

Application on Geographical Locations

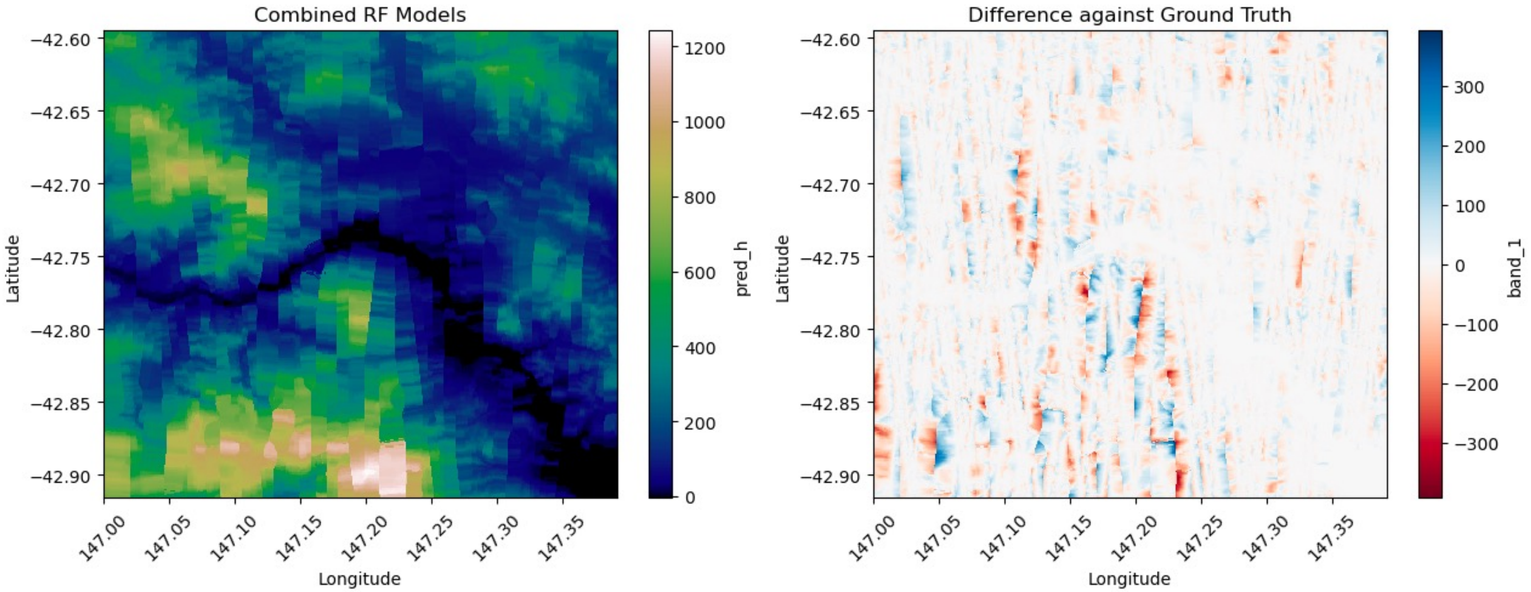


Mt Taranaki	Min Diff	Max Diff	RMSE	MAE
Laplace	-317.31	216.004	**16.11**	6.451
RF (All Features)	-770.826	526.337	**28.424**	11.092

Grand Canyon	Min Diff	Max Diff	RMSE	MAE
Laplace	-792.423	931.947	109.017	69.526
RF (All Features)	-653.547	661.584	76.177	39.448

South Limburg	Min Diff	Max Diff	RMSE	MAE
Laplace	-206.447	65.012	9.544	4.235
RF (All Features)	-120.75	86.09	7.867	3.867

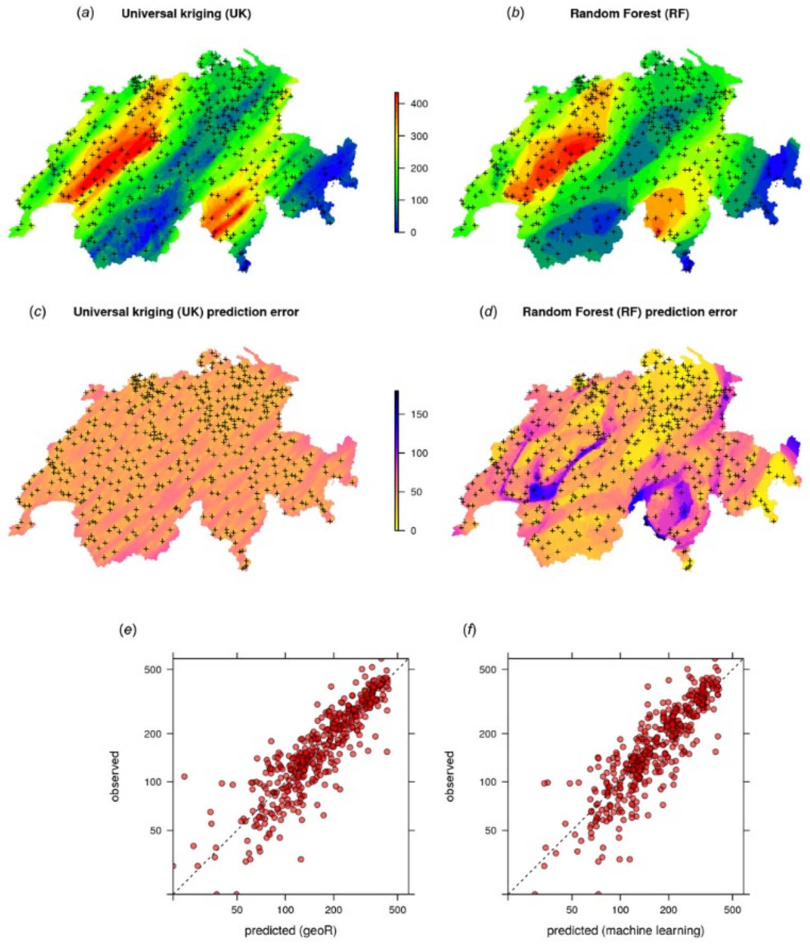
Combined Geographical Locations



Interp’n Method	Min Diff	Max Diff	RMSE	MAE
Laplace	-281.367	169.54	38.543	24.108
RF (Geometric Features)	-342.562	358.449	43.356	24.398
RF (All Features)	-345.907	357.817	43.342	24.384
RF (Combined Models)	-341.878	392.526	43.759	24.606

Conclusion

- **Nearest neighbour height** (importance=0.95) has the most influence/impact on the RF results
- **RF regression** does not always yield better results than Laplace interpolation, depending on terrain. In Taranaki, RF performed worse than Laplace interpolation
- **Hengl et. al. (2018)** - more data crunching algorithms as training dataset does not necessary lead to drastic improvement in accuracy
- **Is it worthwhile to use RF instead of Laplace?** Considering the computation time, and amount of data required to process, it might not be worth



RFSp (Hengl et. al., 2018)

Thank you!