

## ESTIMATING ABOVEGROUND BIOMASS IN ZAGROS FOREST, IRAN, USING SENTINEL-2 DATA

H. Torabzadeh <sup>1,\*</sup>, M. Moradi <sup>2</sup>, P. Fatehi <sup>3</sup>

<sup>1</sup> Engineering faculty, Bu-Ali Sina University, Hamedan, Iran - torabzadeh@basu.ac.ir

<sup>2</sup> Omran Tossee University, Hamedan, Iran - moslem.moradi@gmail.com

<sup>3</sup> Department of Forestry and Forest Economics, University of Tehran, Tehran, Iran - parviz.fatehi@ut.ac.ir

**KEY WORDS:** aboveground biomass (AGB), random forest regression (RFR), Sentinel-2, Zagros forest

### ABSTRACT:

Accurate and reliable assessment of above-ground biomass (AGB) is important for the sustainable forest management, especially in Zagros forests, in which a frangible forest ecosystem is being threatened by anthropogenic factors as well as climate change effects. This study presents a new method for AGB estimation and demonstrates the potential of Sentinel-2 Multi-Spectral Instrument (MSI) data as an alternative to other costly remotely sensed data, such as hyperspectral and LiDAR data in unapproachable regions. Sentinel-2 performance was evaluated for a forest in Kurdistan province, west of Iran, using in-situ measured AGB as a dependent variable and spectral band values and spectral-derived vegetation indices as independent variables in the Random Forest Regression (RFR) algorithm. The influence of the input variables number on AGB prediction was also investigated. The model using all spectral bands plus all derived spectral vegetation indices provided better AGB estimates ( $R^2 = 0.87$  and  $RMSE = 10.75 \text{ t ha}^{-1}$ ). Including the optimal subset of key variables did not improve model variance but slightly reduced the error. This result is explained by the technically-advanced nature of Sentinel-2, which includes fine spatial resolution (10, 20 m) and strategically-positioned bands (red-edge), conducted in different topographical conditions with an advanced machine learning algorithm. However, assessing its transferability to other forest types with varying conditions would enable future performance and interpretability assessments of Sentinel-2.

### 1. INTRODUCTION

Accurate assessment of forest above-ground biomass (AGB) is important for the sustainable management of forests, particularly for Zagros forest areas whose currently degraded through overgrazing and deforestation. In west of Iran, forest is severely influenced by the pressure from human development on vegetated areas, specially landcover change for cultivation and ranching purposes. This, requires timely-based observing of available resource. An assessment of AGB helps foresters and scientists to monitor and understand ecosystem responses (Chinembiri et al., 2013; Gara et al., 2014). Furthermore, time series and frequently monitoring of the forest status provide a basis for decision-making and the sustainable use of forest resources with a view to introducing appropriate planning and conservation efforts.

Traditional field-based methods and remote sensing (RS) methods are known as two major approaches for forest biomass estimation. There is no doubt that traditional methods are more accurate (Lu, 2006), but they are also laborious, difficult to implement in inaccessible areas, time consuming and destructive in nature (Henry et al., 2011). Thus, this study has favoured remote sensing techniques since their inception. AGB cannot be directly measured from space, however, the use of spectrally-derived parameters from sensor-measured reflectance enables increased biomass prediction accuracy when combined with field-based measurements (Dong et al., 2003). Many studies utilized hyperspectral, LiDAR, and medium-resolution sensors with sufficient field data collection to estimate AGB (Chen et al., 2009; Muinonen et al., 2012; Rana et al., 2014;

Dube et al., 2015; Shen et al., 2016). Employing hyperspectral and LiDAR remote sensing technologies, confronts with some restrictions, e.g. high data accusation and processing costs and data redundancy, that have resulted in a shift towards the use of free and readily available broadband, including Landsat and Sentinel-2 (pandit et al., 2018), which offer a large swath width, letting timely AGB estimations from local to regional-scale (Hall et al., 2011; Laurin et al., 2014).

Although, Landsat data have been mostly used for forest AGB estimation (Foody et al., 2003; Powell et al., 2010), its increasing data saturation in fully vegetated areas leads to under-estimation of biomass (Steininger, 2000; Kasischke et al., 2014). Sentinel-2 equipped with a multi-spectral instrument (MSI) sensor, launched on 23 June 2015 by the European Space Agency (ESA), provides a significant improvement in spectral coverage, spatial resolution, and temporal frequency over the current generation of Landsat sensors (Gómez, 2017). It offers a multi-purpose design of 13 spectral bands ranging from visible and near-infrared (NIR) wavelengths to shortwave infra-red wavelengths at 10 m, 20 m and 60 m ground pixel size. Besides, the presence of four bands within the red-edge region, centred at 705 (band 5), 740 (band 6), 783 (band 7), and 865 nm (band 8a) (Shoko and Mutanga, 2017), gives a high potential for mapping various vegetation characteristics, such as vegetation nutrient (Clevers and Gitelson, 2012), assessing rangeland quality (Ramoelo et al., 2015), mapping and monitoring wetlands (Kaplan and Avdan, 2017) and tree canopy cover (Godinho et al., 2017). Sentinel-2 was recently evaluated for forest AGB estimation in tropical forests (Chen et al., 2018; pandit et al., 2018), however, to the best of our knowledge, it

\* Corresponding author

has yet been conducted in the challenging and in-danger Zagros forests. In such areas, Sentinel-2, with its fine resolution and free access policy, can offer new opportunities for timely and accurate AGB estimation.

In this research, Random Forest Regression (RFR) was chosen based on prior studies, which have shown that the RFR approach provides one of the best performances among empirical modelling (Strobl at al., 2008; Karlson at al., 2015). Moreover, RFR is able to determine the relevance of variables, which is crucial for concluding the final AGB estimates based on the incorporated spectral bands and their combinations. Also, ignoring the contribution of irrelevant variables, that can be confusing, results in a faster model and less prone to overfitting (Chave at al., 2005).

The main aim of this study is, therefore, to 1) investigate the performance of spectrally-derived indices using Sentinel-2 MSI combined with in-situ measurements for estimating AGB in the Zagros forests, West of Iran and 2) point the major spectral variable to generate the smallest subset of input variables in the RFR algorithm.

## 2. STUDY AREA AND DATA

### 2.1 Study area

The study area is a relatively highly sloped area, in Kurdistan province, west of Iran, situated in the Northern part of Zagros mountain chain (Figure 1). The forest are mainly dominated by broadleaves species Brant's Oak (*Quercus brantii*) and Aleppo oak (*Quercus infectoria*).

### 2.2 In-situ data

Forest inventory data were collected in 57 forest plots in July 2018, distributed according to a random sampling strategy. All the trees inside each rectangular 20<sup>m</sup> by 20<sup>m</sup> plot were measured. Forest parameters, namely the diameter at breast height (DBH) greater than 5cm and tree height (H), were measured using a tape and a total station, respectively. The species name of each tree was noted as well. Figure 2 shows the conditions in one of the measuring plots.

Among all researches on AGB estimation in Zagros forests, a few studies focused on the Brant's Oak dominated forests, particularly in Kurdistan province (e.g. Abbasi et al. 2017). Therefore, the biomass values were estimated using an existing species-specific allometric equation developed by Yousefvand, et al. (2017), in which similar Brant's Oak dominated forests was investigated. Using the quadratic equation, tree-level AGB values were calculated using equation 1:

$$Y = 0.7116 D^2 - 6.1363 D + 38.473 \quad (1)$$

Where D is tree diameter at breast height (cm) and Y is the tree-based AGB in kilogram. Using the number of trees per plot, the plot-level AGB were standardized in tonnes per hectare (t ha<sup>-1</sup>).



Figure 1. The study area as a black rectangle in (a) and its exact border are presented in a grey polygon.



Figure 2. One of the measuring plots shows Brant's Oak trees locate in a sloped terrain.

### 2.3 Remote sensing data

Single scene standard Sentinel-2 Level-1C product obtained on closest date of acquisition to field collection dates were selected and downloaded. This product is an ortho-image in UTM/WGS 84, 38N projection, with per-pixel radiometric measurements provided for top of atmosphere reflectance. The data is acquired in 13 spectral bands, however, bands 1, 9, and 10 were excluded in this study as they are used to detect atmospheric features. The image was converted from radiance to surface reflectance by applying the dark object subtraction method. Moreover, the 10m spatial resolution bands were resampled to 20-m resolution to unify the remotely sensed data to size of the measured plots.

## 3. METHOD

To test the applicability of the Sentinel-2 data for estimating forest AGB in the study area, spectral bands and a few vegetation indices (VIs), which presented better performance in previous studies (Zhu and Lu, 2015; Chen et al., 2015), were exploited (Table 1).

Estimation of forest AGB using MSI data was based on the extension of a tree-based model called Random Forest Regression (RFR). A detailed description of the basic theory behind the model is provided by Breiman (2001). In this algorithm, decision trees are generated to the maximum extent without pruning using a randomly-selected two thirds of the samples as training data with bootstrapping, which strengthens the flexibility by aggregating the prediction across individual trees to make a final prediction. The rest of the data, i.e., the remaining third, is called OOB data (out-of-bag), is not seen by the model, and is used as validation samples to estimate the model errors (Prasad at al., 2006).

There are two important parameters, namely *mtry*, which is denoted as the number of variables available for splitting at each node of the tree, and *ntree*, which is the number of trees

adjusted to achieve a desirable prediction. These two parameters were optimized to achieve a reasonable prediction (selecting the lowest RMSE). To identify whether a smaller set of the variable would improve model performance, *ntree* and *mtry* were tested in the range of 500 to 1000 and 1 to 20, respectively, which explained the variables optimally. For indicating the importance of variables, a score is assigned, that depends on changes in the error when a particular variable is varied, namely *IncNodePurity*. The larger the effect of a *IncNodePurity*, the more importance is assigned to that variable (Reif et al., 2006). Taking this into account, all the remote sensing-generated variables were used for AGB estimation and the *IncNodePurity* measure was used to determine the variable importance. The RFR was implemented in MATLAB software 2015 version using the statistical toolbox.

In order to evaluate the RFR performance, the coefficient of determination ( $R^2$ ) and root mean square error (RMSE) between the estimated AGB and the field-measured AGB, were calculated.

Spectral feature	description
MSI bands:	
B2	Blue
B3	Green
B4	Red
B5	RE 1
B6	RE 2
B7	RE 3
B8	NIR
B8a	RE 4
B11	SWIR1
B12	SWIR2
Vegetation indices:	
NDVI	$(NIR - R) / (NIR + R)$
RGR	$Red_{665} / Green_{560}$
EVI	$2.5 * ((NIR - R) / (1 + NIR + 6R - 7.5Blue))$
SR	$NIR / RED$
PSRI	$(\rho_{665} - \rho_{560} / \rho_{740})$
SAVI	$(NIR - R) / (NIR + R + L) * 1.5$
Red-edge-based NDVIs:	
1	$(NIR - RE1) / (NIR + RE1)$
2	$(NIR - RE2) / (NIR + RE2)$
3	$(NIR - RE3) / (NIR + RE3)$
4	$(NIR - RE4) / (NIR + RE4)$

Table 1. Spectral bands and calculated vegetation indices from Sentinel-2 MSI.

#### 4. RESULTS AND DISCUSSION

Forest stand parameters (DBH and H) measured for individual trees within the rectangular plot were aggregated to generate plot-level AGB for all sampling plots in the study area. The average AGB in the area is 35.70 t  $hec^{-1}$  with 18.61 t  $hec^{-1}$  standard deviation, while the lowest and highest AGB are 52.44 t  $hec^{-1}$  and 21.09 t  $hec^{-1}$ , respectively. These results are close to destructive measurements have done before in the study area (Abbasi et al., 2017).

The parameters *ntrees* and *mtry* was adjusted to generate a better prediction result. Boosting of *ntree* for the full predictor variable resulted in a value of 700, whereas *mtry* was 14, which produced the lowest RMSE.

Considering the potential of the RFR algorithm for predicting forest biomass, the RFR model with all input variables ( $n = 20$ ) produced a reasonable result ( $R^2 = 0.87$  and  $RMSE = 10.75$  t  $ha^{-1}$ ).

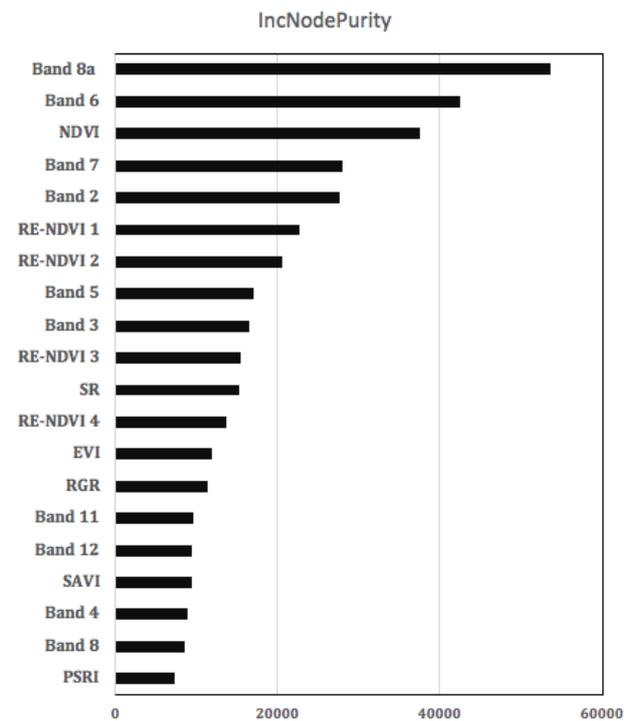


Figure 3. Relative importance of the variables

RFR also calculates the relevance of the input variables using *IncNodePurity*. As shown in Figure 3, the most influential features pose in the red edge spectral region, which confirm the results achieved in pandit et al., (2018). The spectral band 8a in the most important feature, however, band 8 in the same spectral region belongs to the impractical features. It may related to the spectral resolution, while band 8a has five times lower bandwidth. The variable selection method used in this study identified the smallest subset of predictor variables in the model.

No. of Variables Used	Eliminated Variable (Backward)	$R^2$	RMSE
20	Full variables	0.87	10.57
19	PSRI	0.87	10.55
18	Band 8	0.86	10.01
17	Band 4	0.87	10.08
16	SAVI	0.86	10.69
15	Band 12	0.84	11.01
14	Band 11	0.83	10.93
13	RGR	0.83	11.42
12	EVI	0.81	11.80
11	RE-NDVI 4	0.80	12.76
10	SR	0.80	12.95
9	RE-NDVI 3	0.80	13.33
8	Band 3	0.79	13.82
7	Band 5	0.78	14.00
6	RE-NDVI 2	0.78	16.07
5	RE-NDVI 1	0.74	19.46
4	Band 2	0.74	20.04
3	Band 7	0.73	22.58
2	NDVI	0.69	26.99

1	Band 6	0.61	31.14
---	--------	------	-------

Table 2. Performance of the RFR by withdrawing of variables.

Comparing the variable importance values (*IncNodePurity*), progressive backward feature elimination did not substantially improve model performance. Removal of the five least important variables from the model resulted no change in  $R^2$ ; however, as presented in Table 2, there were inconsistencies in the RMSE value. Slight increases or decreases were observed when the variables were progressively removed from the model. Thus, the full set of predictor variables that yielded the highest  $R^2$  (0.87) and low RMSE (10.08 t ha<sup>-1</sup>) values were considered in the final RFR algorithm to predict the AGB of the forest. Finally, an AGB map was produced using the best predictor variables from the final model generated by the RFR algorithm (Figure 4). The choice of Sentinel-2 spectral bands and spectral-derived VIs for producing biomass estimates was based on the fact that it produced strongly-explained variable values ( $R^2$ ) and a low RMSE. The average predicted forest biomass was 37.45 t ha<sup>-1</sup>, ranging from 18.70 t ha<sup>-1</sup> to 65.63 t ha<sup>-1</sup>.

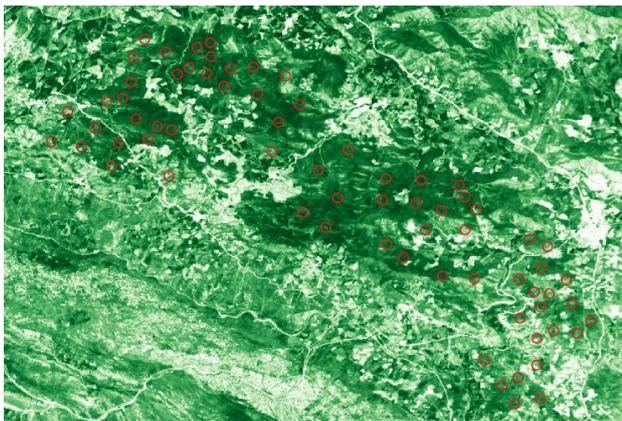


Figure 4. Estimated AGB using best predictor from RFR algorithm, where dark green indicates maximum AGB (65.63 t ha<sup>-1</sup>) and light green indicates lowest value (37.45 t ha<sup>-1</sup>). Red circles point to the measuring plots.

## 5. CONCLUSIONS

This study investigates the performance of the RF algorithm in predicting forest AGB in Zagros forests, West of Iran, using fine spatial resolution Sentinel-2 MSI data. The approached results show Sentinel-2 data effectively predicted the above-ground biomass of the Zagros forests, with an  $R^2$  value of 0.87 and an RMSE value of 10.75 t ha<sup>-1</sup>. Also, selection of important variables did not improve the variance explained by the RFR model ( $R^2$ ), but improved the model performance error (RMSE) from 10.75 to 10.08 t ha<sup>-1</sup>.

Technical improvements in the Sentinel-2 MSI sensor at fine medium resolution (10 and 20 m) have the potential to enable accurate predictions of AGB in areas of semi-arid forest with sloped terrain.

Since Sentinel-2 MSI is a relatively new sensor, in comparison to Landsat and MODIS, we recommend further researches in near future, such as the applicability of this data to other parts of the forested Zagros mountain and including species-specific allometric estimates for plot-level AGB values.

## REFERENCES

- Abbasi, L.; Shakeri, Z.; Shabaniyan, N. and Moreno, G. Branch and leaf biomass of Lebanon oak (*Quercus libani* Oliv.) and gall oak (*Q. infectoria* Oliv.) trees in different years after pollarding. *Iranian Journal of Forest and Poplar Research*, 2017, Vol. 25 No. 1.
- Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32.
- Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree allometry and improved estimation of carbon stocks and balance of tropical forest. *Ecosyst. Ecol.* 2005, 145, 87–99.
- Chen, J.; Gu, S.; Shen, M.; Tang, Y.; Matsushita, B. Estimating aboveground biomass of grassland having a high canopy cover; an exploratory analysis of in situ hyperspectral data. *Int. J. Remote Sens.* 2009, 30, 6497–6515.
- Chen, L., Ren, C., Zhang, B., Wang, Z. and Xi, Y., 2018. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with Sentinel imagery. *Forests*, 9(10), p.582.
- Chinembiri, T.S.; Bronsveld, M.C.; Rossiter, D.G.; Dube, T. The precision of C stock estimation in the Ludhikola watershed using model-based and design-based approaches. *Nat. Res. Res.* 2013, 22, 297–309.
- Clevers, J.G.P.W.; Gitelson, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* 2012, 23, 344–351.
- Dong, J.; Kaufmann, R.K.; Myneni, R.B.; Tucker, C.J.; Kauppi, P.E.; Liski, J.; Buermann, W.; Alexeyev, V.; Hughes, M.K. Remote Sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks. *Remote Sens. Environ.* 2003, 84, 393–410.
- Dube, T.; Mutanga, O.; Elhadi, A.; Ismail, R. Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution space borne multispectral RapidEye sensor and advance machine learning algorithms. *Remote Sens.* 2014, 14, 15348–15370.
- Foody, G.M.; Boyd, D.S.; Cutler, M.E. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* 2003, 85, 463–474.
- Gara, T.; Murwira, A.; Chivhenge, E.; Dube, T.; Bangira, T. Estimating wood volume from canopy area in deciduous woodlands of Zimbabwe. *South. For.* 2014, 76, 237–244.
- Godinho, S.; Guiomar, N.; Gil, A. Estimating tree canopy percentage in Mediterranean silvopastoral systems in using Sentinel-2A imagery and the stochastic gradient boosting algorithm. *Int. J. Remote Sens.* 2017, 1–23.
- Gómez, M.G.C. Joint Use of Sentinel-1 and Sentinel-2 for Land Cover Classification: A Machine Learning Approach. Master's. Thesis, Lund University, Lund, Sweden, 2017.
- Hall, F.G.; Bergen, K.; Blair, J.B.; Dubayah, R.; Houghton, R.; Hurr, G.; Kellndorfer, J.; Lefsky, M.; Ranson, J.; Saatchi, S.; et al. Characterizing 3D vegetation structure from space; mission requirements. *Remote Sens. Environ.* 2011, 115, 2753–2775.
- Henry, M., Picard, N., Trotta, C., Manlay, R., Valentini, R., Bernoux, M. and Saint André, L., 2011. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. *Silva Fennica*, 45(3B), pp.477–569.
- Kaplan, G.; Avdan, U. Mapping and monitoring wetlands using Sentinel-2 satellite imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 2017, IV-4/W4, 271–277.
- Karlson, M.; Ostwald, M.; Reese, H.; Sanou, J.; Tankoano, B.; Mattsson, E. Mapping tree canopy cover and aboveground

- biomass in Sudano-Sahelian Woodlands using Landsat 8 and Random forest. *Remote Sens.* 2015, 7, 10017–10041.
- Kasischke, E.S.; Goetz, S.; Hansen, M.C.; Ozdogan, M.; Rogan, J.; Ustin, S.L.; Woodcock, C.E. *Remote Sensing for Natural Resource Management and Environmental Monitoring*, 3rd ed.; John and Wiley and Sons, Inc.: Hoboken, NJ, USA, 2014.
- Laurin, G.V.; Chen, Q.; Lindsell, J.A.; Coomes, D.A.; Del Frate, F.; Guerriero, L.; Pirotti, F.; Valentini, R. Aboveground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* 2014, 89, 49–58.
- Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. *International journal of remote sensing*, 27(7), pp.1297-1328.
- Muinonen, E.; Parikka, H.; Pokharel, Y.P.; Shrestha, S.M.; Eerikäinen, K. Utilizing a multi-source forest inventory technique, MODIS data and Landsat TM images in the production of forest cover and volume maps for the Terai Physiographic Zone in Nepal. *Remote Sens.* 2012, 4, 3920–3947.
- Pandit, S., Tsuyuki, S. and Dube, T., 2018. Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. *Remote Sensing*, 10(4), p.601.
- Powell, S.L.; Cohen, W.B.; Healey, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Ohmann, J.L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens. Environ.* 2010, 114, 1053–1068.
- Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 2006, 9, 181–199.
- Ramoelo, A.; Cho, M.; Mathieu, R.; Skidmore, A.K. Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl. Remote Sens. Environ.* 2015, 124, 516–533.
- Rana, P.; Korhonen, L.; Gautam, B.; Tokola, T. Effect of field plot location on estimating tropical forest above ground biomass in Nepal using airborne laser scanning data. *ISPRS J. Photogramm. Remote Sens.* 2014, 94, 55–62.
- Shen, W.; Li, M.; Huang, C.; Wei, A. Quantifying live aboveground biomass and forest disturbance of mountainous natural and plantation forests in Northern Guangdong, China, based on Multi-temporal Landsat, PALSAR and field plot data. *Remote Sens.* 2016, 8, 595.
- Shoko, C.; Mutanga, O. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS J. Photogramm. Remote Sens.* 2017, 129, 32–40.
- Steininger, M.K. Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *Int. J. Remote Sens.* 2000, 21, 1139–1157.
- Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* 2008, 9, 307.
- Zhu, X. and Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, pp.222-231.
- Yousofvand, M.; Soosani, J.; Ostakh, E.; and Hosseinzadeh, L., 2017. Estimate the above ground biomass in Brant's oak (*Quercus brantii* Lindl.). *J. of Wood & Forest Science and Technology*, Vol. 24 (4).