



Vienna University of Technology
Faculty of Mathematics and Geoinformation
Department of Geodesy and Geoinformation

MASTERTHESIS

An approach for developing a fuel moisture content dataset for Austrian wildfire danger assessment

submitted in fulfillment of the requirements for the degree of
Diplom-Ingenieur

as part of the Master's program
Geodesy and Geoinformation

by
Christian Johannes Plakolm BSc
Matriculation number 01629058

at the
Department of Geodesy and Geoinformation
of the **Faculty of Mathematics and Geoinformation**
of the **Technische Universität Wien**

under the supervision of
Univ.Prof. Dr.rer.nat. Wouter Arnoud Dorigo and
Projektass. Piet Emanuel Büechi MSc and
Univ.Ass.in Dipl.-Ing.in Ruxandra-Maria Zotta

Vienna, in April 2024

Signature author

Signature supervisor



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Declaration

I hereby declare that I have written this thesis by myself without any help or assistance of others. External literature used to clarify the content or provided data sources are fully cited. All mentioned information is in accordance with fact or truth up to my knowledge.

Vienna, in June 2024

Abstract

Fuel moisture content (FMC) of live vegetation is an important factor for wildfire ignition and propagation. Therefore, for wildfire danger assessment, it is inevitable to consider this parameter in models regarding the prediction of the ignition probability and spreading behaviour of wildfires. A current prototype of an integrated forest fire danger assessment system in Austria is introducing FMC as a functional derivative of meteorological conditions, which was calibrated for Canadian pine trees. To improve wildfire danger assessment in Austria, an updated FMC dataset that considers the local environmental and vegetational conditions is required. Observations from optical remote sensing instruments are sensitive to FMC in the infrared domain and thus, offering the possibility to estimate this quantity on a broader scale.

An established global dataset for FMC is derived from the Moderate-Resolution Imaging Spectroradiometer (MODIS), which delivers temporally frequent reflectance data with a spatial resolution of 500 m. Modelling FMC with this pixel size might be too coarse when considering narrow valleys or mountain peaks. In this study, satellite observations from the Sentinel-2 mission were used to calculate a new FMC dataset for the area of Styria with a high spatial resolution of 20 m. The aim was to develop a workflow that should further allow the FMC estimation for the entire Austrian country, which can support the domestic wildfire danger assessment. This was done by applying a pretrained machine learning model to Styrian Sentinel-2 data. This model was trained on Australian Sentinel-2 observations that were related to the MODIS derived FMC values. The validation of this approach with MODIS FMC data of Styria led to rather low correlations (ρ between 0 and 0.16) between the two datasets.

A main reason for the low correlation results was identified as the relevant climatic and vegetational differences between Australia and the study area of Styria. Another reason was the different origin of the data. While the Sentinel-2 reflectances of Australia were subjected to a preprocessing algorithm, the Styrian data was originally taken from the Google Earth

Engine tool. To develop a valid FMC dataset for Austria based on optical remote sensing data and machine learning algorithms, the training of a new model is recommended for further research work. The inclusion of a data layer that contains locally optimized FMC estimations with high spatial resolution remains a desirable goal for the Austrian wildfire danger assessment.

Contents

1	Introduction	1
1.1	Objective of this thesis	3
2	Fuel Moisture Content and its role in wildfire events	4
2.1	What is Fuel Moisture Content?	4
2.2	Methods of FMC estimation	5
2.2.1	Field measurements	5
2.2.2	Meteorological models	5
2.2.3	Satellite-based FMC estimation	6
2.3	Integration of FMC in wildfire danger assessment	10
3	Study area	12
4	Materials and methods	14
4.1	Australian machine learning model	14
4.2	Sentinel-2 observations	16
4.3	MODIS Fuel Moisture Content Product	18
4.4	Austrian Wildfire Database	19
4.5	Applying the Random forest model to Styrian Sentinel-2 data	20
4.6	Analyzing the applicability of the random forest model	21
4.7	Validation of FMC estimates with MODIS FMC data	24
4.7.1	Sampling and interpolation of FMC datasets	24
4.7.2	Methods for correlation analysis	26
4.8	Validation of FMC estimates with wildfire events	28
5	Results	30
5.1	Estimation of FMC with Sentinel-2 observations	30
5.2	Area of applicability (AOA) of the random forest model	32

5.3	Correlation between Sentinel-2 FMC estimations and the MODIS FMC data product	34
5.3.1	Correlation between single observation days	34
5.3.2	Temporal correlation	38
5.4	Evaluation of the interrelation between wildfire events and FMC	39
6	Discussion	42
6.1	Deriving FMC from Sentinel-2 data	42
6.2	Applicability of the random forest model	43
6.3	Validation with MODIS FMC data	45
6.4	Validation with wildfire data	47
6.5	Potential improvements	48
7	Conclusion and outlook	50
A	Observation days of used Sentinel-2 and MODIS data	51
	List of Figures	53
	List of Tables	55
	Bibliography	56

1 Introduction

As climate change is moving forward, the number and severity of forest fires is increasing in many parts of the world [Müller et al., 2020b]. Also in Austria a trend has been detected that more and larger forest fires are occurring due to warmer and dryer weather conditions [Müller et al., 2015]. A recent evidence for this development has been brought in 2021, as in Lower Austria the most devastating wildfire of the country took place in Hirschwang/Rax where about 115 ha burnt down and emergency costs of about 30 million euros were caused [Dzugan, 2022]. The reasons for the ignitions of those wildfire events may vary, as there are anthropogenic as well as natural causes like lightnings, which are the major roots of wildfire events [Müller and Vacik, 2017].

In addition to analyzing wildfire data, implementing precautionary measures, adapted forest management, raising awareness and collecting empirical data on fire behavior, wildfire danger assessment is a crucial task for the authorities in order to mitigate such severe fire impacts and to minimize suppression costs like for the case of Hirschwang/Rax [Müller et al., 2020b; De Angelis et al., 2015]. Currently known approaches for that are often based on meteorological variables [De Angelis et al., 2015]. For example, Kalabokidis et al. [2013] have developed a web-based GIS platform, which integrates real-time weather data to help fire management professionals with the deployment planning.

However, there are multiple environmental conditions that favour wildfire ignition and propagation and must therefore be taken into account for holistic fire danger assessment. For the European Alps, Müller et al. [2020b] have developed a prototype called Integrated forest Fire Danger assessment System (IFDS), which considers fire causes, vegetation and topography additionally to fire weather. This IFDS prototype was designed for the case study region of Austria. A specific data layer, which contributes to the vegetational factor in wildfire modeling, was assigned to the fuel moisture of live vegetation, which proved to be a crucial parameter for forest fire events in previous research work [Wang et al., 2019; Yebra et al., 2013; García et al., 2020; Quan et al., 2021b].

Fuel moisture content (FMC) is describing the amount of water in plants as a ratio of the water mass and the dry weight of the plant [Yebra et al., 2013]. This quantity is relevant for wildfire prediction and modeling as vegetation with lower water content is more likely to catch fire, which is affecting wildfire ignition and the rate and direction of its spread [Rossa et al., 2016]. This is described more detailed in Chapter 2. However, a quantification of the direct relation from FMC to forest fire events is yet to be fully understood in science [Rao et al., 2023].

Also the estimation of FMC on a larger scale has been a challenge in past research efforts as field measurements are labour-intensive and limited in its validity when representing larger areas. Nevertheless, their local accuracy is an important benefit for the calibration and validation of remotely sensed FMC data [Yebra et al., 2013]. There have been many different approaches in the past for modeling FMC based on satellite observations or meteorological measurements, which are explained in Chapter 2.2, but many of them are limited in their applicability for Austrian wildfire management because their parameters are locally calibrated with in-situ FMC data or other local measurements from different countries [Van Wagner et al., 1987; Yebra et al., 2018].

As an example of satellite-based FMC estimation, Quan et al. [2021b] have invented a global FMC dataset with 500 m spatial and 1-day temporal resolution depending on data of the Moderate Resolution Imaging Spectroradiometer (MODIS), which was validated with field measurements from 8 worldwide spread countries, which led to overall good correlation results. Larraondo [2021] has built up on this FMC dataset to develop a machine learning model for estimating FMC in Australia from Sentinel-2 satellite observations, which offers the potential to even improve the spatial resolution to 20 m resolution with a temporal resolution of 2 to 3 days for mid-latitudes European Space Agency [2015]. For the conceptual IFDS approach by Müller et al. [2020b], fuel moisture data from the Canadian Fire Weather Index was taken into account, where FMC is modelled as a function of meteorological conditions affecting the fuel moisture of Canadian pine trees [Van Wagner et al., 1987].

So, to give an outline for this thesis, a lack of available FMC data that is optimized for Austrian environmental conditions leaves much room for the improvement of current methods of wildfire prediction in Austria and therefore new approaches for FMC dataset generation may have to be tested to achieve this goal.

1.1 Objective of this thesis

The objective of this thesis is to generate a new FMC dataset creation workflow for the territory of Styria that is locally validated and can be further used to improve wildfire danger assessment systems for the entire country of Austria. This dataset should be capable of predicting FMC with high spatial and temporal resolution. This was approached by applying the method of Larraondo [2021] to the optical remote sensing data from the Sentinel-2 satellites. Furthermore, it should be evaluated if this method is sufficient for estimating reliable FMC values for the environmental conditions of Styria.

2 Fuel Moisture Content and its role in wildfire events

2.1 What is Fuel Moisture Content?

Fuel Moisture Content (FMC) is defined as the ratio between the mass of water in a plant and its respective dry weight. In equation 2.1 this ratio is described as the difference of the weight from a fresh plant sample (m_f) and its dry mass (m_d) [Yebra et al., 2013]:

$$FMC = \frac{m_f - m_d}{m_d} \quad (2.1)$$

This direct relation is used for the measurement of FMC from in-situ samples, which are weighed after extraction and reweighed after they were oven-dried [Yebra et al., 2019]. For the determination of FMC with satellite-based data, a different physical definition is more appropriate since the accurate estimation of the dry mass of vegetation is more challenging. This is because of less specific absorption features from dry matter than from the amount of water [Fourty and Baret, 1997; Riano et al., 2005]. Therefore, the canopy water content (CWC) as the product of the equivalent water thickness (EWT) and the leaf area index (LAI) is used as a related quantity to derive fuel moisture from remote sensing methods [Yebra et al., 2013]. The EWT is calculated as the amount of leaf water divided by its respective leaf area A [Riano et al., 2005]:

$$EWT = \frac{m_f - m_d}{A} \quad (2.2)$$

$$CWC = EWT * LAI \quad (2.3)$$

The CWC is defined as the amount of canopy water per unit ground area and given in g/m^2 [Yebra et al., 2013]. This quantity is likely to be estimated more accurately from remote

sensing techniques than FMC [Bowyer and Danson, 2004] and has been related to it in previous research work [Yebra et al., 2013].

2.2 Methods of FMC estimation

Current methods and developed datasets of FMC estimation can be separated into field measurements, meteorological models and satellite-based techniques.

2.2.1 Field measurements

As mentioned in Chapter 2.1, field measurements of FMC are based on the gravimetric determination of the fresh and oven-dried mass of a representative sample of small twigs and leaves. These in-situ measurements are further used for the calibration and validation of satellited-based FMC estimation methods and therefore some requirements in terms of data quality have to be fulfilled. The challenge is to collect a sample that takes the much coarser resolution of remote sensing data into account. Since field samples are usually taken from areas sized between 0.01 to 0.1 ha and satellite data used for FMC estimation is scaled to pixel sizes from 0.1 to 100 ha, field measurements of FMC should aim for low spatial variations in their sampling sites. Therefore, those sites should be as homogeneous as possible and not dominated by different vegetation or tree species over an area of 100 ha [Yebra et al., 2013].

Yebra et al. [2019] have collected those worldwide measurements and built up a global database called *Globe-LFMC*, which covers in-situ data of FMC from 1977 to 2018 gathered from 1,383 sampling sites in 11 countries.

2.2.2 Meteorological models

To estimate FMC based on meteorological conditions, models have been developed in the past. Viney [1992] has reviewed a couple of models that are related to the moisture content of fine, dead fuels. These models are mainly based on the parameters dry-bulb temperature and relative humidity.

Byram [1943], for example, has invented a functional relationship between fuel temperature and air temperature dependent on the downward flux of solar radiation and the wind speed. The fuel and air temperature were also used in combination with the relative air humidity

to model the relative humidity of fuels. Later on, the equilibrium moisture content of wood and forest litter was in focus of laboratory experiments to derive physical models based on air temperature and humidity. Van Wagner [1972] has modeled this equilibrium moisture content for adsorption (E_w , obtained by wetting from below) and desorption (E_d , obtained by drying from above) with following equations:

$$E_w = 0.942H^{0.679} + 0.000499e^{0.1H} + 0.18(21.1 - T)(1 - e^{-0.115H}) \quad (2.4)$$

$$E_d = 0.618H^{0.753} + 0.000454e^{0.1H} + 0.18(21.1 - T)(1 - e^{-0.115H}) \quad (2.5)$$

This approach - among others - is the basis of vapour exchange modeling. Van Wagner et al. [1987] have created the Fine Fuel Moisture Code based on experiments on different types of pine trees. The moisture content calculation is related to the moisture content of the day before (m_0) and is different for a wetting (m_w) or drying period (m_d):

$$m_w = E_w - (E_w - m_0) * 10^{-k_w} \quad (2.6)$$

$$m_d = E_d + (m_0 - E_d) * 10^{-k_d} \quad (2.7)$$

The k_w or k_d is the appropriate wetting rate or log drying. If m_0 is in between E_w and E_d then no change in moisture content is assumed and therefore $m = m_0$ [Van Wagner et al., 1987]. The additional effect of precipitation on fuel moisture content has also been modeled in the past by different scientists [Viney, 1992] and is also considered in the Fine Fuel Moisture Code by Van Wagner et al. [1987] with additional terms on equations 2.6 and 2.7.

2.2.3 Satellite-based FMC estimation

There have been several approaches for estimating FMC on a broader scale with satellite data in previous research work. The use of remote sensing data for FMC estimation is enabled as especially the near and short-wave infrared spectral regions are sensitive to variations in FMC [Bowyer and Danson, 2004]. This fact is used in estimation models, which can be either statistical or physical-based [Yebra et al., 2013].

Physical-based FMC models

Physical models have played an important role in FMC estimation in the past as they relate reflectance data directly to canopy variables by applying physical laws, which allows their

use for larger areas [Quan et al., 2021b]. Yebra et al. [2018] have created a FMC dataset on a continental scale as they applied a radiative transfer model (RTM) inversion to Australian optical remote sensing data from the MODIS instrument, which lead to a result of $R^2 = 0.58$ when validating the model with field measurements. Quan et al. [2021b] extended this research to a global scale, verified with field measurements from 8 countries, and achieved an even better result ($R^2 = 0.71$). Also to active microwave satellite observations, physical models for FMC estimation can be applied. Wang et al. [2019] calibrated and validated a model for dual-polarimetric Sentinel-1A backscatter data with FMC field measurements from a sampling site in Texas/USA, leading to a dataset with high spatial resolution (5×20 m) and a temporal resolution of 12 days, which outperformed the FMC estimation from optical Landsat-8 data after cross-validation in terms of the root mean square error.

Those physical approaches based on RTMs are characterized with a high degree of complexity as the selection and parametrization of the model is a challenging task, especially when heterogeneous environmental conditions are present in the application. Uncertainties may also occur because of the ill-posed problem, where similar simulated spectra of the RTM are caused by very different input parameters [Yebra et al., 2013].

Statistical FMC models

Statistical models are fitting a functional relationship between spectral reflectance data and field measurements of FMC, but are limited as they are site-specific and not applicable to larger scales if those sites are not covering a broad range of leaf and canopy characteristics [Yebra et al., 2013].

García et al. [2020] have produced a FMC dataset from optical remote sensing data by using a statistical approach. They related the reflectance data of the Landsat-5 TM satellite to multiple FMC field measurements of 9 sites in Western USA by fitting linear regression models to specific spectral indices. The Enhanced Vegetation Index performed best among the tested indices with a correlation coefficient of 0.44, which could be improved to 0.69 when normalizing the spectral indices and FMC field data. Nevertheless, the limitation of site-specificity of this approach was shown with a significant increase in correlation up to over 0.9 when only taking single sites into account for model fitting.

Forkel et al. [2022] have also tested a statistical model for global FMC estimation by relating passive microwave satellite observations of vegetation optical depth to the MODIS based

FMC dataset of Yebra et al. [2018] and the field measurements of Yebra et al. [2019]. With a logistic regression model, it was possible to achieve medium to high correlations in grass- and shrubland, but averagely weak correlations for forested areas. The spatial resolution of this FMC data product is with $0.25^\circ \times 0.25^\circ$ much coarser than the MODIS FMC dataset ($0.005^\circ \times 0.005^\circ$).

When statistical models are based on multiple input variables or complex physical models are applied to remote sensing data, machine learning workflows are a helpful tool to use. For land cover mapping, convolutional neural networks and random forest models have been tested [Boston et al., 2022], and also the application of RTMs on optical remote sensing data has been used by Quan et al. [2021a] to estimate foliage fuel load, another critical variable in wildfire prediction.

This mentioned random forest algorithm relies on multiple decision trees with random complexity [Ho, 1995]. According to the initial concept by Ho [1995], each one of these trees is constructed in a subset of the feature space of the dataset and therefore develops its own prediction schema. This approach of using multiple predictors leads to high accuracy on the unseen test data. The basic principle was later extended, a major focus in subsequent research was on the construction of random decision trees with the aim to increase their independence and reduce correlations between them [Breiman, 1996; Ho, 1998; Breiman, 2001]. According to those studies, a bootstrap aggregating algorithm is used to split the training data into several samples, where one data point can occur in more than one sample. For each training set a classification and regression tree is constructed in a subset of the feature space. All trees together are building the ensemble called random forest [Sun et al., 2024]. This is shown schematically in Figure 1.

Larraondo [2021] has tried out a random forest model for a regression task, the estimation of FMC, which is described in particular in section 4.1. In this case, the decision trees take on numerical values and the prediction result is the average value over all predictor trees. This also allows calculating statistical measurements like standard deviations over the single tree predictions for validating the model [Breiman, 2001].

The statistical model of Larraondo was further used in this thesis for FMC estimation, as the application of such an empiric model is suitable for immediate and easy-to-implement solutions. The complexity of physical models would require much more effort for the

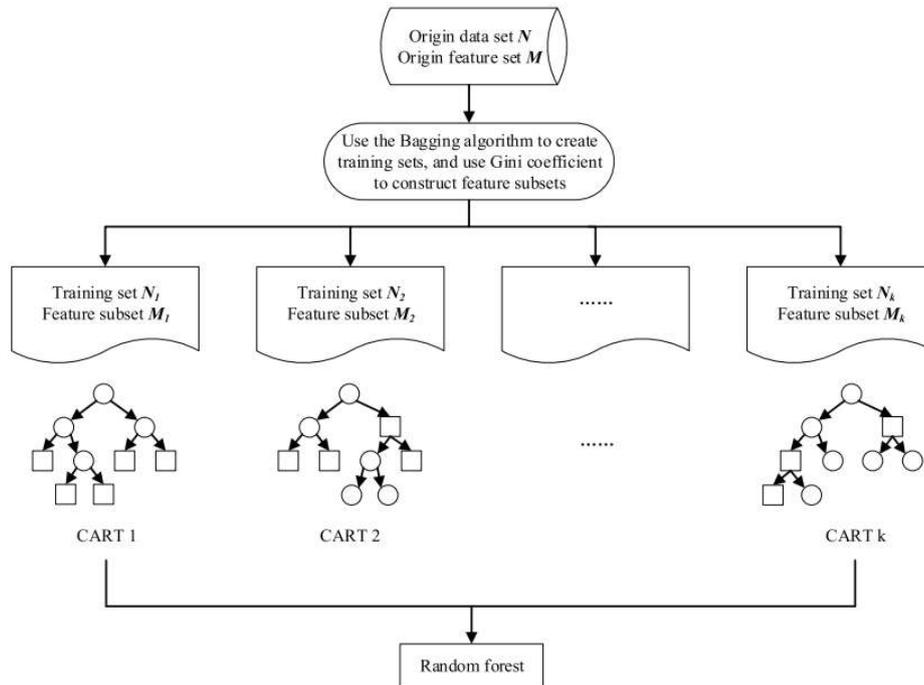


Figure 1: Construction process of the random forest, built up by classification and regression trees (CARTs) [Sun et al., 2024]

parametrization of a new model and high experience with the physical characteristics of the specific remote sensing instrument that is used. The site-specificity of statistical models remains as a weakness compared to physical models, while the ill-posed problem is a source for uncertainties when working with the latter [Yebra et al., 2013].

A meteorological model has the disadvantage that it is not directly related to measures of vegetation water content [Van Wagner et al., 1987]. This is limiting the applicability of those models in alpine areas during the winter season and early spring, where the snow on south-facing slopes is melting earlier due their higher sun exposure. This increases the risk of a forest fire in those areas in spring [Müller et al., 2020b]. This effect is not covered by weather indices, as they are not designed for a use in cold seasons [Van Wagner et al., 1987]. Therefore, a meteorological approach for FMC estimation was not pursued further in this study.

2.3 Integration of FMC in wildfire danger assessment

FMC is directly influencing the occurrence and behaviour of wildfires, as plants with higher FMC are delaying the fire ignition and the water content in live vegetation has a cooling effect, which reduces the rate of spread [Rossa et al., 2016; Yebra et al., 2019]. Thus, for effective fire danger assessment, it is inevitable to take this parameter into account when estimating wildfire danger.

Yebra et al. [2018] have used their MODIS based FMC dataset to create a flammability index determined with logistic regression models and training on occurred fire events. This offered the perspective to predict wildfire risk with adequate accuracy as the "Area under the Curve" values from a ROC (receiver operating characteristic) plot were between 0.7 and 0.8 for the different vegetation types. For the modeling of wildfires in the Mediterranean area, Kondylatos et al. [2023] have created a large-scale dataset containing variables of important wildfire drivers that can be used in machine learning algorithms. FMC is not explicitly integrated there, the normalized difference vegetation index (NDVI) and the LAI derived from MODIS are related variables covered in this dataset.

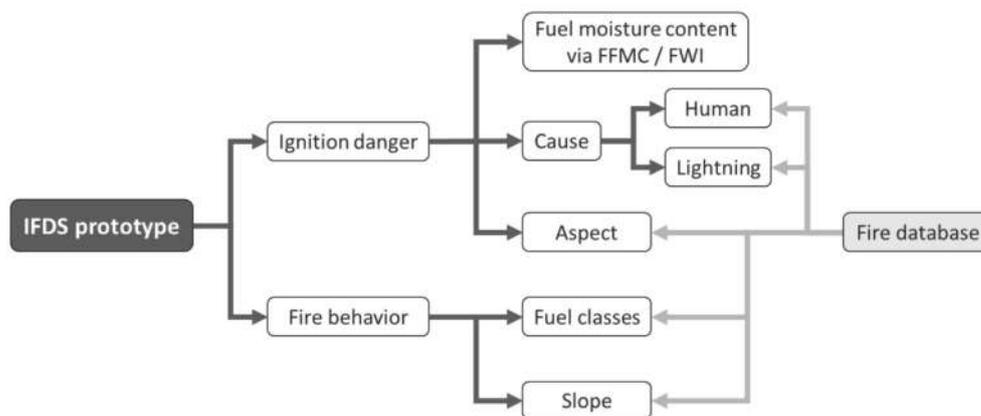


Figure 2: Included data layers in the prototype of the integrated fire danger assessment system [Müller et al., 2020b]

For the European Alps, Müller et al. [2020b] have created a prototype of an integrated fire danger assessment system introduced for the territory of Austria within the research project CONFIRM (Copernicus Data for Novel High-resolution Wildfire Danger Services in Mountain Regions). In Figure 2 the considered variables are illustrated, FMC was included

with the Fine Fuel Moisture Code of the Canadian Fire Weather Index, mentioned in Chapter 2.2.2. This dataset is not calculated with satellite observations but with meteorological input parameters that are affecting FMC. The functional relationship of these variables and fuel moisture is calibrated for Canadian pine trees and consequently not designed for the application in Austrian forests [Van Wagner et al., 1987]. Nevertheless, the validation with forest fire data of Austria from 2018 and 2019 resulted in better overall prediction accuracy than those of common fire weather indices, but still with room for improvements of this framework [Müller et al., 2020b].

3 Study area

As study area, the Austrian federal state Styria was selected as it is known for its high proportion of forest (61.4%), which has earned it the nickname "the green heart of Austria" [Breitfelder et al., 2019]. An impression of the vegetational structure in Styria is given in Figure 3, where the land cover map is depicted including the district capitals.

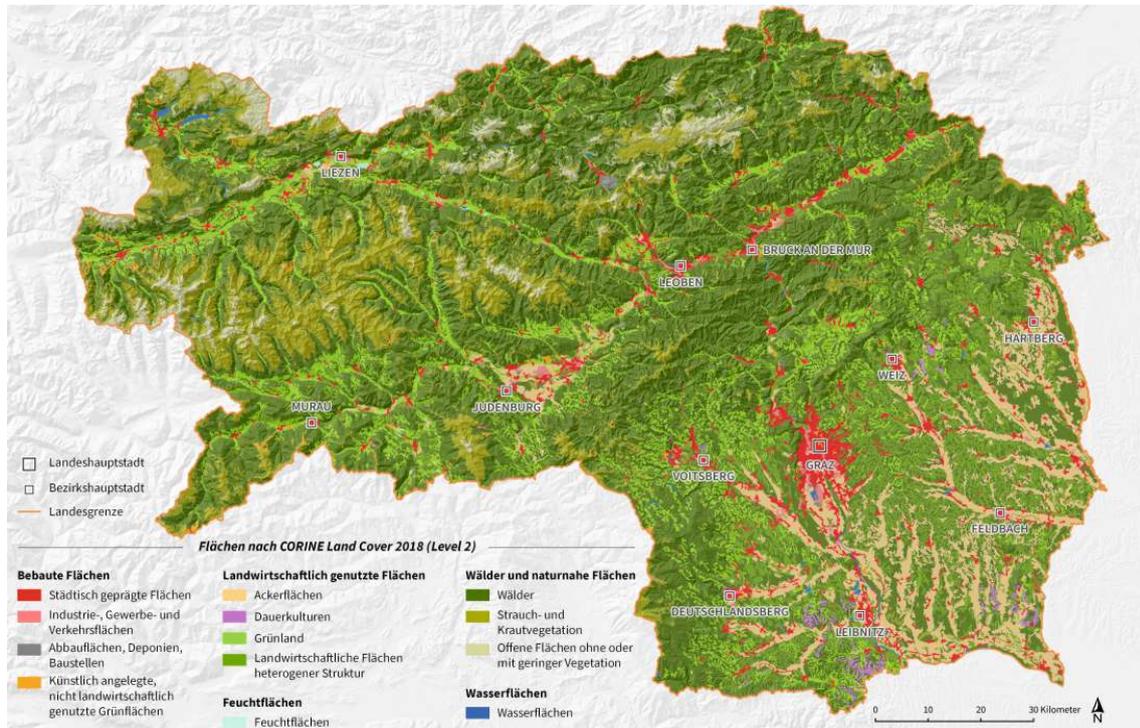


Figure 3: Land cover map of Styria [Schulatlas Steiermark, 2018]

The forested areas are especially dominating in the mountainous areas in the northern and western part of Styria. Common tree species are spruce and fir at higher elevations, in the lower regions beeches and other warmth-loving deciduous tree types are favoured [Breitfelder et al., 2019]. As elevation increases, the vegetation changes to shrubs and herbs.

A few open areas of the main alpine ridge can be found, where the elevation is exceeding the vegetation line and therefore no significant planting exists. The valleys in the mountainous part of the federal state are dominated by grassland, while the highest amount of cropland can be found in the valleys of the southern part.

Climatologically, Styria can be divided roughly into two regions. In the northern part the alpine climate prevails, while the southern part is categorized into the climate of the alpine foothills. The latter one includes moderately cool winters with little snow and warm summers, winters with little snow and warm summers, often combined with thunderstorms and hail. In contrast to that, the alpine climate is characterized by lower temperature and higher precipitation as the elevation increases. The temperature difference between the mountain peaks and the southern valleys is around 14 °C [Breitfelder et al., 2019].

This difference is also captured by the Köppen-Geiger classification, which indicates that the lower parts of the country have a warm temperate and fully humid climate with warm summers (class Cfb). In contrast to that, the mountainous regions are classified as areas with a humid snow climate with warm summer months (Dfb) or even cool ones (Dfc), the higher the elevation gets [Kottek et al., 2006].

4 Materials and methods

For developing a new Fuel Moisture Content dataset with high spatial and temporal resolution, a research method was built up. In the following subchapters, the selected datasets and existing models are described as well as their way of use in this thesis.

4.1 Australian machine learning model

Recently, Larraondo [2021] developed a new model for estimating FMC from multi-spectral imaging satellites. This model is leading to a major increase in spatial resolution of FMC datasets as it is relying on observations of the Sentinel-2 satellites with its ground pixel size of 10 to 60 meters [European Space Agency, 2015]. So, for the purpose of generating a FMC dataset with high spatial resolution, this approach for FMC estimation is preferred for investigation. The model is based on a machine learning workflow, precisely on the random forest algorithm (see Chapter 2.2.3).

For training and testing the random forest model by Larraondo [2021], Australian FMC and Sentinel-2 datasets were used. Reflectance values and spectral indices derived from Sentinel-2 observations are related to FMC estimates in Australia derived from the Moderate-Resolution Imaging Spectroradiometer (MODIS). This collection of FMC estimates covers the period from July 2015 until the end of 2018 and was provided for the training and testing dataset with a temporal frequency of four days. Additionally, some masking and filtering workflows have been established to improve validity and quality of the MODIS data. In detail, a quality mask has been created, which saves information of the vegetation type in the research site, whether it is grassland, shrubland or forest. Areas where the vegetation type classification is not constant over time between 2015 and 2018 are masked from further processing. Furthermore, only sites where FMC values have been registered on at least 90% of the observation days in the research period have been added to the final dataset. The resulting masked raster has been filtered again with a 3x3 kernel to only accept data points where the 3x3 surrounding area is compound of all non-masked

pixels with identical vegetation type. The coordinates of the remaining 2,999 data points, which were used for training and testing the random forest model are visualized in Figure 4.

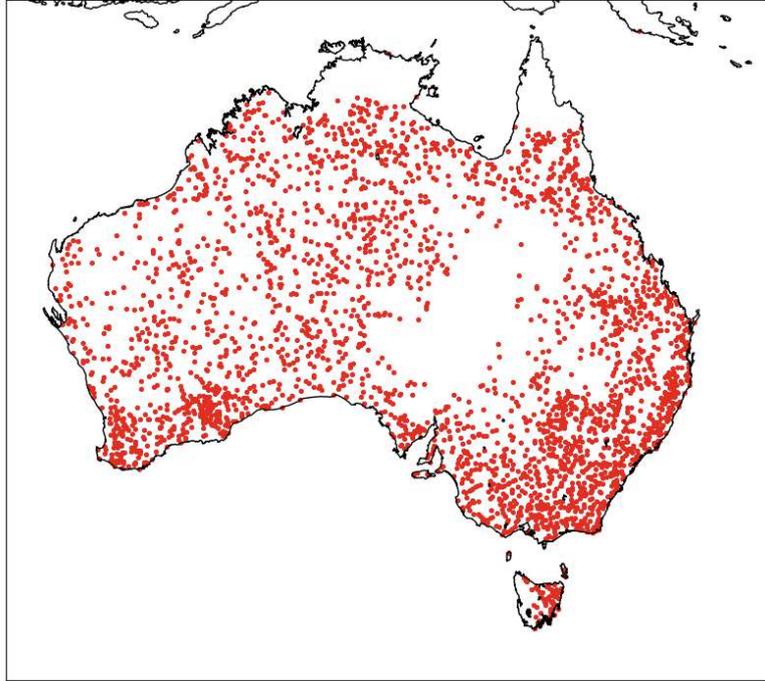


Figure 4: FMC data points used for training the random forest model

Those locations deliver in total 873,374 FMC estimations, whose distribution is depicted in Figure 5. It can be seen that a majority of the FMC values lies within the range 0 to 120% with a peak in the range from about 50 to 65% [Larraondo, 2021].

The coordinates of these data points, known with their latitude and longitude, was then used to extract Sentinel-2 observations from the Open Data Cube of *Digital Earth Australia*. The observations were extracted in a spatial range from $\pm 0.001^\circ$ in latitude and longitude around the data point coordinates and the reflectance values of the model-relevant Sentinel-2 bands were averaged in the extracted area for every timestamp. These reflectance values were resampled and interpolated to the specific dates of the extracted MODIS FMC estimates [Larraondo, 2021].

With the resulting combined dataset of Sentinel-2 reflectances and MODIS FMC estimations, the random forest model was trained [Larraondo, 2021]. According to Larraondo, only 15% of

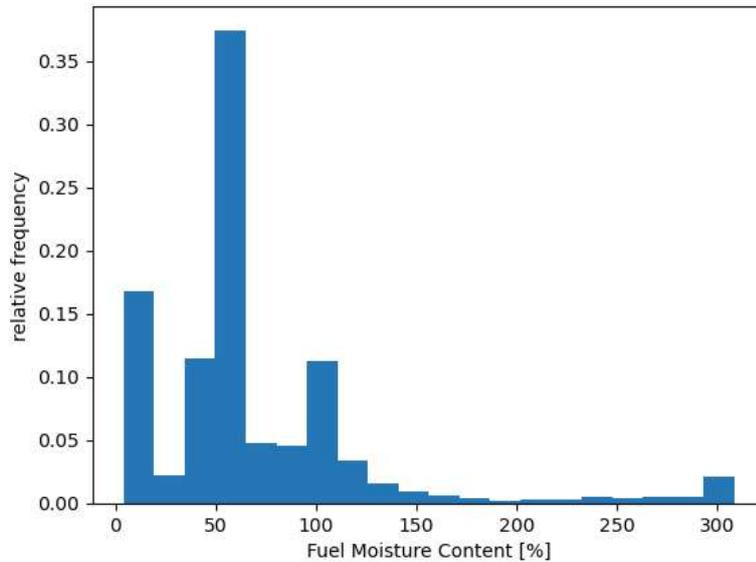


Figure 5: Distribution of FMC values in the training data

the 873,374 data entries were taken over all vegetation types to be split in training data (75%) for model generation and test data (25%) for model validation. The *RandomForestRegressor* module from the Python library *scikit-learn* [Pedregosa et al., 2011] was used to train the machine learning model. In the model of Larraondo, the number of decision trees was set to 25, where each tree has a maximum depth of 25, which limits the expansion of nodes for each decision tree. These parameter values were selected by hyperparameter tuning, where tree numbers of 10, 25 and 50 were combined with tree depths of 5, 10, 25 and 50. As the criterion for splitting the training data along the trees' nodes, the mean squared error was selected. The generated model was used to predict FMC values for the test data to analyze its performance. Overall, the coefficient of determination (R^2) between predicted and real FMC values was 0.53. For grassland, the prediction fitted the best ($R^2 = 0.70$), while for shrubland ($R^2 = 0.41$) and forest ($R^2 = 0.47$), the correlations were lower [Larraondo, 2021]. The importance of the individual features for FMC estimation is shown in Figure 6 and will be used for further analysis.

4.2 Sentinel-2 observations

To estimate FMC values for Styria by using the random forest model of Larraondo [2021], Sentinel-2 observations are needed as the input data of the model. Sentinel-2 is a 2015 launched satellite mission contributing to the Copernicus programme of the European Space

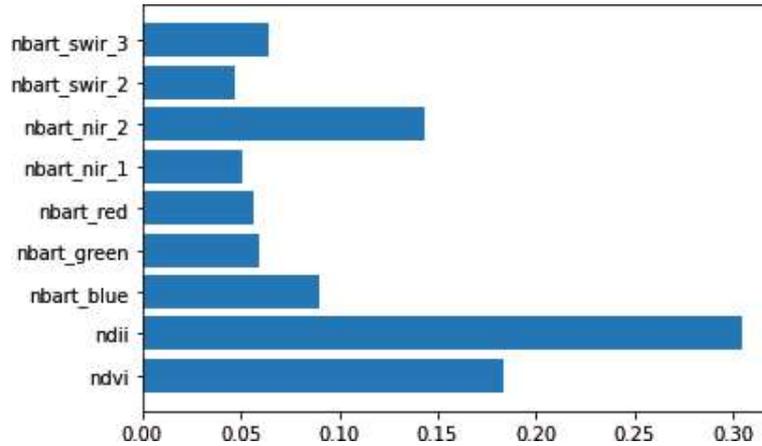


Figure 6: feature importance for estimating FMC with the Australian random forest model

Agency [European Space Agency, 2015]. Two multi-spectral imaging satellites (Sentinel-2A and Sentinel-2B) in the same sun-synchronous orbit but shifted by 180° guarantee a revisit time of every 5 days at the equator and every 2 to 3 days in mid-latitudes like the Styrian research area. The Sentinel-2 mission covers 13 bands in the visible, near infrared and short-wave infrared segment of the electromagnetic spectrum [Drusch et al., 2012].

For this study, Sentinel-2 images were accessed via Google Earth Engine. The images are available as atmospherically corrected and scene classified measures of surface reflectance (Level-2A) [European Space Agency, 2023]. For the estimation of FMC, the reflectance data of the three optical bands (blue (B2), green (B3), red (B4)) and each two bands in the near (B8 & B8A) and short-wave infrared (B11 & B12) spectrum was used [Larraondo, 2021]. The optical bands and the first near infrared band (with smaller wavelength) offer a spatial resolution of 10 m, the second infrared band (with longer wavelength) and the short-wave infrared bands are accessible at a pixel size of 20 m [European Space Agency, 2023]. As Tiede et al. [2021] had good results in deriving more or less cloud-free Sentinel-2 images in the alpine region with it, a cloud cover maximum of 20% was set to filter the extracted images. Afterwards, the data from the quality indicating band QA60 (described from European Space Agency [2023] as cloud mask), with a spatial resolution of 60 m, was used to remove pixels affected by the rest of the clouds and cirrus. Additionally, the Scene Classification Map, which is also part of every Sentinel-2 Level-2A image [European Space Agency, 2023], was analyzed to delete pixels of snow that would not lead to rational FMC

estimates. The resulting filtered images were clipped to the territory of Styria and saved as GeoTIFF files like presented as RGB composite in Figure 7.



Figure 7: RGB composite of a filtered Sentinel-2 image of Styria

Furthermore, two spectral indices are used as input for the random forest model of Larraondo, the normalized difference vegetation index (NDVI) and the normalized difference infrared index (NDII). These indices are calculated as the following [Cucca et al., 2020]:

$$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \quad (4.1)$$

$$NDII = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \quad (4.2)$$

ρ_{NIR} , ρ_{SWIR} and ρ_R represent the reflectance values of the near infrared (B8), shortwave infrared (B11) and red (B4) band of the Sentinel-2 observations.

The NDVI and NDII together with the seven bands from the Sentinel-2 observations are building up the nine-dimensional predictor space of the random forest model [Larraondo, 2021].

4.3 MODIS Fuel Moisture Content Product

With the input of Sentinel-2 reflectance data, the random forest model was trained to predict the MODIS FMC estimates in selected Australian sites [Larraondo, 2021]. The MODIS instrument is mounted on board of the Terra and Aqua satellite of the National

Aeronautics and Space Administration (NASA). These two satellites are imaging the whole Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands with spatial resolutions of 250, 500 or 1000 meters [National Aeronautics and Space Administration, 2023].

The MODIS data product MCD43A4 (Collection 6) provides daily Nadir Bidirectional Reflectance Distribution Function-Adjusted reflectance data with a spatial resolution of 500 meters using 16-day composites of MODIS data. This data is supplied with simplified mandatory quality layers for MODIS band 1 through 7 [Land Processes Distributed Active Archive Center, 2023]. Quan et al. [2021b] have invented a workflow to estimate daily FMC values on a global scale using this MODIS data product, as well as the MODIS land cover product MCD12Q1 for separating different fuel classes. The estimation of FMC was done with radiative transfer models (RTMs) inversion techniques, where the MCD15A3H product, containing the LAI, was used for parameterizing the RTM. The resulting dataset was validated with 3,034 in-situ measurements from 120 worldwide sites located in Australia, China, Italy, Senegal, South Africa, Spain, Tunisia and the USA. The accuracy of the MODIS FMC estimations compared to those field measurements - after anomalous FMC measurements were filtered out - was quantified with $R^2 = 0.71$ and a root-mean-square error of 32.36% (with a p-value of < 0.01) [Quan et al., 2021b].

In this study, this global dataset is not only fundamental for the training of the random forest model from Larraondo with Australian data, but also used as a FMC reference for the validation of FMC estimations in Styria. Due to storage limitations, the FMC dataset derived from MODIS was not provided with daily estimations, but with a temporal resolution of 8 days given in WGS84 coordinates for this research [Quan et al., 2021b].

4.4 Austrian Wildfire Database

For investigation of the correlation between FMC estimations and fire ignition in Styria, the Austrian Wildfire Database delivers information about wildfire events that occurred in the past (accessible via <http://fire.boku.ac.at>). This database was created by the Institute of Silviculture from the University of Natural Resources and Life Sciences Vienna [Vacik, 2023]. For the generation of this directory, data of major fire events was collected from 1993 on to the present. The information provided includes date, location, burned area, cause of the fire, affected vegetation and tree species, fire type, fire behavior, number of action forces, helicopters and fire brigades involved [Müller et al., 2020a]. This information can be

accessed via a Web GIS interface. For the purpose of this study, forest fire data of Styria in 2019 and 2021 was used. These periods were selected based on the amount of available, nearly cloud-free Sentinel-2 data for comparison.

4.5 Applying the Random forest model to Styrian Sentinel-2 data

To generate FMC datasets for Styria with Sentinel-2 images, observation days had to be found where a major part of the Styrian territory was depicted without exceeding the determined cloud cover limit of 20%. Especially in the western part of the federal state, which is dominated by the main alpine ridge, Sentinel-2 image tiles were filtered out very often because of too cloudy conditions. The covered area of Sentinel-2 observations was also limited by the footprints of the swaths from the two satellites, which are not overlapping in general on subsequent orbits around the earth [European Space Agency, 2016]. On some days a single swath of a Sentinel-2 satellite covered the whole territory of Styria, but sometimes the space between two subsequent swaths caused that only strips of the western and the eastern part of the research area were observed on subsequent days. Such images were combined to one GeoTIFF indexed with the observation date of the first swath. The consideration of all observation days with a sufficient Styrian area covered - not only those with full coverage - should enable a more comprehensive evaluation of temporal dynamics and effects in FMC.

For the analysis of time series, two periods were chosen based on the amount of available data. A shorter period during June and July 2021 with 9 observation days was selected for initial evaluations of the validity of the model and of the ability to depict the temporal dynamics of FMC. For monitoring the FMC estimation throughout a longer period with different environmental conditions, 26 Sentinel-2 datasets registered from January to December 2019 were used. The respective observation dates can be gathered from Table 2.

For all further processing steps, *Python 3.7.12* has been used [Van Rossum and Drake, 2009]. To import the individual GeoTIFF files for all listed observation dates, the Python library *Rasterio* was applied [Gillies et al., 2013]. With this library, the data of the different bands could be extracted to separate spatial 2D-arrays, which is furthermore needed for the input of the random forest model and the calculation of the NDVI and NDII. All the bands were provided from Google Earth Engine with a pixel size of 20 m. Therefore, the individual

bands have a uniform spatial resolution since the bands in the visible spectrum have been resampled to 20 m. As the Sentinel-2 reflectance data accessed from Google Earth Engine is scaled by a factor of 0.0001, all values had to be multiplied by 10000 to fit them to the model requirements [European Space Agency, 2023]. As the random forest model is not able to process NaN-values [Larraondo, 2021], all the pixels that were masked with NaN due to cloud cover, cirrus or snow were given a value of zero among all bands. To set their FMC estimations back to NaN after prediction, a mask had been created beforehand, which distinguishes between masked (0) and non-masked (1) pixels. For fulfilling the input requirements of the random forest model, a 2D-array with the dimensions $n \times 9$ has been stacked containing the reflectance data and the calculated spectral indices from the Sentinel-2 image. This array is built up, so that for each of n image pixels a row is assigned containing the Sentinel-2 data in the following order: NDVI, NDII, B4, B3, B2, B8, B8A, B11, B12 [Larraondo, 2021]. All these processing steps have been done with the Python library *numpy* [Harris et al., 2020].

The 2D-array was handed to the random forest model as input for FMC prediction. The random forest model has been loaded to the Python-script as a file of the Python datatype *pickle* [Van Rossum and Drake, 2009; Larraondo, 2021]. Used with the *scikit-learn* library [Pedregosa et al., 2011], the model predicted FMC values for the input data [Larraondo, 2021]. After reshaping the $n \times 1$ output array back to its spatial dimensions and setting the FMC value for masked pixels to NaN, the output could be visualized with the Python library *matplotlib* and the corresponding array was saved to a txt-file for further analysis and validation [Harris et al., 2020; Hunter, 2007]. An example of FMC estimation from a Sentinel-2 image is shown in Figure 8.

4.6 Analyzing the applicability of the random forest model

As the random forest model was trained on Australian FMC samples, a major question of this research is if this model can be applied on Styrian Sentinel-2 in a meaningful way, so the predicted FMC values are reliable. As mentioned in Chapter 2.2.3, ensemble regressors like random forest algorithms allow a straightforward calculation of uncertainty measures due to their combination of multiple predictors. The variation in predictions among individual trees can be used to quantify the uncertainty of the model prediction, but as this is based on the knowledge of the model, it does not address the difference in environmental conditions between the training area and the application area [Meyer and Pebesma, 2021]. As this

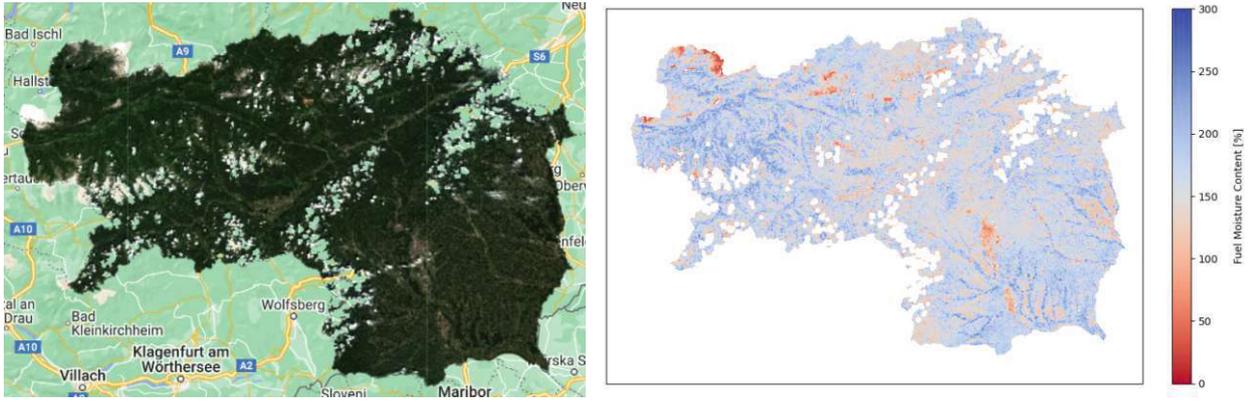


Figure 8: Estimation of FMC (right) from a Sentinel-2 image (left as RGB composite) of 29/07/2021

difference might be significant between Australia and Styria, a quantifier of the dissimilarity between training and new data is needed.

Meyer and Pebesma [2021] have suggested a methodology to calculate the area of applicability (AOA) of a machine learning model, which is used in this study. This area is defined as the domain in the predictor space where the model is able to learn about correlations based on the training data and the mentioned uncertainty quantification of the random forest model holds. For this purpose, the 9 predictor variables - consisting of the reflectance of 7 Sentinel-2 bands and the calculated NDVI and NDII - have to be standardized by calculating the difference of the respective observations to the corresponding mean value and dividing it through the standard deviation of the predictor variable (see equation 4.3).

$$X_{i,j}^s = (X_{i,j} - \bar{X}_j) / \sigma_j \quad (4.3)$$

$X_{i,j}^s$ refers to the standardized value of the i -th observation of the j -th predictor variable, while \bar{X}_j and σ_j stands for the arithmetic mean and standard deviation of the j -th predictor variable, calculated over the whole training data.

Because not all variables have the same effect on FMC prediction, these standardized reflectance observations have to be weighted regarding their influence on the model. The feature importance values displayed in Figure 6 serve as weights w_j for equation 4.4:

$$X_{i,j}^{sw} = X_{i,j}^s * w_j \quad (4.4)$$

For determining the AOA, the dissimilarity in the training data itself is the key factor. To have a measure for that, the Euclidean distances between all data points are calculated like the following example for two random points a and b :

$$d(a, b) = \sqrt{\sum_{j=1}^9 (X_{a,j}^{sw} - X_{b,j}^{sw})^2} \quad (4.5)$$

For further computations, the mean value \bar{d} of all these distances in the training data is important. For a new data point k in Styria, the distance d_k to the nearest training data point is used to calculate the Dissimilarity Index (DI). Before the calculation of d_k , the Styrian Sentinel-2 data has to be weighted and standardized to the mean and standard deviation of the training data variables, analogous to equations 4.3 and 4.4. The Dissimilarity Index DI_k of such a new data point is then defined as the ratio between the minimum distance to the training data space and the average of all pairwise distances in the training dataset:

$$DI_k = d_k / \bar{d} \quad (4.6)$$

The DI can take on values from 0 to ∞ , 0 if the new data point is identical to a training data point. If the DI exceeds a value of 1, the distance to the nearest data point is greater than the average dissimilarity in between the training data space.

Finally, the AOA is defined as the area where the DI does not surpass a specific threshold, which is dependent of the training data properties. This threshold is derived from the DI values of the training data points. As the DI values should not be 0, the indices of the training data points are calculated in relation to a different sample of the training dataset, which is used for the cross-validation of the model. The outlier-removed maximum DI value of the training data is then used as the threshold to decide if a new data point is within the AOA. The outlier removal is done with setting the limit to the sum of the 75-percentile and 1.5 times the interquartil range of the determined DI values. This fulfills the definition of the AOA, that the machine learning model can be applied to areas, which are similar to areas used for cross-validating the trained model [Meyer and Pebesma, 2021].

For reasons of computational effort, only 1% of the training data was randomly selected for the calculation of the AOA threshold (Note: The used machine learning model is trained with

11.25% of the dataset.). 25% of this subsample are separated as cross-validation fold. This leads to 8,734 data points - registered from July 2015 to December 2018 and distributed over the entire Australian country - which are used for the calculation of \bar{d} . 2,184 points are then excluded for the determination of the DI threshold. To investigate the dissimilarity between Styrian Sentinel-2 data and a subsample of the Australian training data, the calculations mentioned above were applied on three Sentinel-2 images of Styria taken in 2019, each one in spring (21/04), summer (20/07) and fall (23/10). The reflectance values and spectral indices of those observations were each related to a different random subset of the training data for calculating the DI and furthermore the AOA. This method is serving the possibility to detect seasonal and spatial changes of the dissimilarity between Styrian and Australian Sentinel-2 data. The DI threshold for the AOA was calculated independently six times with a random 1%-sample and the average of the six results rounded to two decimal places was taken as a uniform threshold, which enhances the reliability of this value.

4.7 Validation of FMC estimates with MODIS FMC data

To validate the FMC estimation from Sentinel-2 images, the MODIS FMC dataset described in 4.3 has been used as a reference. As the resolutions of these two datasets are different, they had to be sampled to the same pixel size. For this purpose, two options were tested, which are explained in Chapter 4.7.1. After the resolutions were identical, the random forest estimations were compared to the MODIS FMC values for the whole research area. This has been done in two different ways: a static correlation analysis of two datasets with almost equal timestamps and a temporal correlation analysis, which investigates if changes in FMC had been monitored similarly from both datasets. This is part of Chapter 4.7.2.

4.7.1 Sampling and interpolation of FMC datasets

To unify the spatial resolutions of both FMC datasets, two approaches have been made. On the one hand, the MODIS FMC dataset with 500 m resolution was upsampled to 20 m pixel size. On the other hand, the FMC estimations from Sentinel-2 data were downsampled from 20 m to 500 m resolutions. Thus, one sampled dataset could be compared with the original counterpart.

The MODIS datasets were provided as netCDF-files for each relevant year and contained latitude and longitude coordinates in WGS84 and the corresponding FMC value as well

as the respective Unix timestamps [Quan et al., 2021b]. This data was read out with the Python library *netCDF4* [Whitaker et al., 2019]. To upsample the MODIS FMC data of one day to the Sentinel-2 grid, latitude and longitude had to be converted to UTM 33N coordinates in x and y . This was done with the Python library *pyproj* [Snow et al., 2021]. After cutting the MODIS dataset to the research area, the FMC values were interpolated to the coordinates of the Sentinel-2 grid with the function *interpolate.griddata* from the Python library *scipy* [Virtanen et al., 2020]. As interpolation method, the option *linear* was taken, where the input data is triangulated and a linear barycentric interpolation is performed on each triangle. Areas that have been masked from the MODIS dataset were masked again after upsampling by using a k-d tree to look for the nearest MODIS data point to a Sentinel-2 grid point and mask this pixel if the distance to the nearest neighbor is over 500 m [Virtanen et al., 2020], which matches the MODIS FMC resolution. The transformation of the 500 m grid to the 20 m grid is shown exemplary in Figure 9.

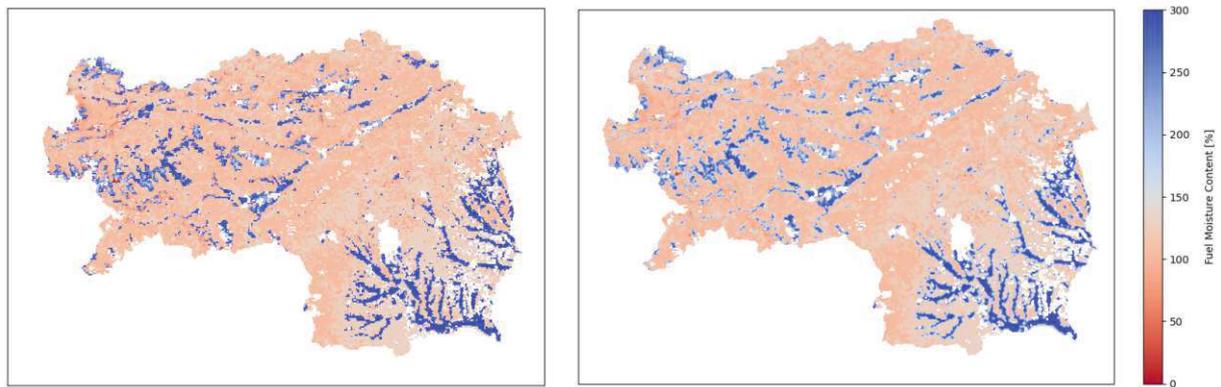


Figure 9: Upsampling of a original 500 m MODIS FMC dataset (left) to the Sentinel-2 grid of 20 m resolution (right)

As it may occur that the FMC estimation with 20 m resolution does not fit the MODIS dataset due to additional uncertainties in the upsampled MODIS FMC product, another approach for validation has been made. The high-resolution FMC estimations from Sentinel-2 were downsampled to the MODIS grid of 500 m resolution, which reduces noise in our FMC approximations. This was done similarly to the MODIS FMC upsampling in terms of the interpolation method. The coordinate transformation was done vice versa to latitude and longitude, therefore the distance threshold for masking areas without processed Sentinel-2 data had to be determined in degrees. It was set to 0.002° as this was

approximately the spacing of the transformed Sentinel-2 grid in latitude and longitude. The downsampling process is exemplary shown in Figure 10.

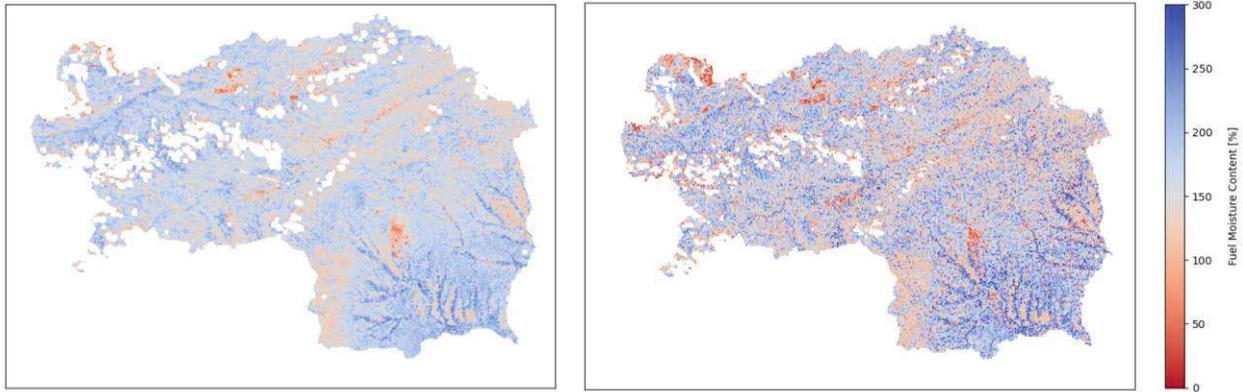


Figure 10: Downsampling of the Sentinel-2 FMC estimations with 20 m pixel size(left) to the MODIS FMC grid of 500 m resolution (right)

This interpolated dataset can then be validated with the original MODIS FMC product. The comparison of the correlations from the two different sampling techniques allows to have an insight, if creating a valid FMC dataset with 20 m pixel size is too optimistic in terms of spatial resolution because of possibly better correlation results between the 500 m datasets.

4.7.2 Methods for correlation analysis

To analyze the correlation between the estimation of FMC based on Sentinel-2 observations and the MODIS FMC product, firstly, a static analysis was carried out both for the 20 m and the 500 m resolution. To the date of the individual Sentinel-2 FMC estimation, the nearest date possible from the 8-day-separated MODIS FMC datasets was chosen for validation. As the MODIS FMC calculations are based on reflectance data composed of 16 days around the index date, there was no great importance attached to this uncertainty. The FMC values from pixels with the same coordinates of one original and one resampled dataset were used for computing the Pearson correlation coefficient ρ . Pixels that have been masked in either one or both of the datasets were excluded from the correlation analysis. ρ is then calculated as the following [Benesty et al., 2008]:

$$\rho_{S2,MODIS} = \frac{cov[S2, MODIS]}{\sigma_{S2}\sigma_{MODIS}} \quad (4.7)$$

$cov[S2, MODIS]$ is the covariance between the Sentinel-2 and MODIS FMC derivatives and σ_{S2} and σ_{MODIS} are the standard deviations of those datasets. The result ρ describes the linear dependency between those two datasets [Benesty et al., 2008]. The square of the Pearson correlation coefficient is the coefficient of determination R^2 . This is a measure of the amount of variance in the Sentinel-2 FMC estimations explained by the MODIS FMC dataset [Laerd Statistics, 2020]. Its value is between 0 and 1. If the two datasets are uncorrelated, R^2 would be 0. The higher the value of R^2 , the bigger is the linear dependency between the two datasets [Benesty et al., 2008]. As the performance of the random forest model during training and testing was also quantified with R^2 [Larraondo, 2021], this value allows a comparison between the model testing on Australian data and the application on Styrian data.

ρ and R^2 were also calculated only for the pixels inside the AOA (see Chapter 4.6) to see if correlation can be improved with omitting data points that are highly dissimilar to the training data. As a detailed investigation of the correlation coefficients, which were computed with data of the whole Styrian research area (so also with data points outside the AOA), the local autocorrelation was computed. This was done by applying formula 4.7 to a 15x15 moving window over the FMC estimation dataset. Thus, every data pixel provided ρ for the 300 x 300 m area around it and spatial differences in the correlation between the Sentinel-2 FMC estimations and the MODIS FMC data could be detected. These methods of correlation analysis were applied to the corresponding FMC estimations of the three Sentinel-2 observation datasets that have been used in the AOA determination (21/04/2019, 20/07/2019, 23/10/2019).

Not only the correlation between an individual Sentinel-2 FMC estimation and a corresponding MODIS FMC dataset was of interest in this study, but also the question if the FMC estimation of the random forest model can detect temporal changes in FMC similarly to the MODIS FMC dataset. To answer this question, all 26 Sentinel-2 FMC estimations from 2019 were compared to 24 MODIS FMC datasets of the same year. The timestamps of the two FMC time series are listed in Table 2. The individual MODIS FMC datasets were covering more or less the whole Styrian research area, except for masked pixels due to snow, clouds and areas that did not fit into the land cover types grassland, shrubland or forest (e.g. urban areas, sparsely vegetated areas) [Yebra et al., 2018]. For

processing those time series data, dataframes of the Python library *pandas* were used [McKinney, 2010; The pandas development team, 2020].

For computing the correlation between two time series, their timestamps have to be identical [Zhang et al., 2003]. Therefore, the Sentinel-2 FMC estimations and the MODIS FMC values were interpolated to the timestamps of the opposite time series with a linear fit between the temporal adjacent FMC values. Extrapolations on the margins of the time series were not permitted [Virtanen et al., 2020]. In total, 48 points in time of the year 2019 were used to analyze the temporal correlation between both datasets. For every individual pixel, the correlation between the two FMC time series was computed by deriving ρ and R^2 . As some pixels may have been masked more often during the used 2019 data or the Sentinel-2 observations were not covering specific areas of Styria on the individual registered days, a threshold for the minimum number of FMC estimations was applied for calculating meaningful correlation quantities based on the available data. If a pixel contained less than 11 (of maximum 26) Sentinel-2 FMC estimations or less than 16 (of maximum 24) MODIS FMC values, it was masked from the correlation analysis. These thresholds were determined by counting the FMC datasets with nearly full coverage of the research area.

4.8 Validation of FMC estimates with wildfire events

The Austrian Wildfire Database (Chapter 4.4) was another reference in the validation of the Sentinel-2 based FMC estimations. One wildfire in Styria in July 2021 was considered for a initial visualization of the relation between MODIS FMC and the Sentinel-2 FMC estimations for the shorter data period in 2021 (see Table 2) and their association with a wildfire event. As a more extensive investigation, all 50 documented wildfires in Styria in 2019 [Vacik, 2023] were taken into account without distinguishing between the wildfires in terms of cause and the size of the burned area.

The locations of those wildfire events were gathered from the database and transmitted to the Sentinel-2 FMC estimations and the upsampled MODIS FMC dataset with 20 m resolution. A 5x5 window was formed around the locations to compute the mean FMC value of the 100 times 100 m area surrounding the burned area for every dataset in the Sentinel-2 and MODIS FMC time series of 2019 (for 50 wildfire events) and 2021 (for 1 wildfire event). The trends in FMC before and after a wildfire for individual events were -

at first - visually evaluated from the corresponding graphs of Sentinel-2 and MODIS derived FMC.

As an overall investigation of the relation between all 50 Styrian wildfires in 2019 and the FMC time series, it was determined if the wildfires occurred to a time with lower FMC in this area. Therefore, the mean FMC values in the 100x100 m surrounding areas on the specific wildfire dates were linearly interpolated from the time series. Then, the quantiles of the FMC distribution in those areas throughout the year were calculated from the time series data. This was done by not taking the mean FMC values into account, but all 25 FMC values in the 100x100 m area of every timestamp. In this way, a bigger and further more reliable amount of data was available for quantile calculation. The quantiles of the lowest 10%, 25% and 50% were determined for all wildfire locations. The interpolated mean FMC value of the 100x100 m area around the wildfire was then compared to the quantiles and assigned to the matching interval. This was done for the Sentinel-2 and the MODIS derived FMC time series. With this analysis, it can be researched if there is a clear dependency between occurred wildfires and low FMC for one of the datasets and how different these evaluations are for the two FMC datasets.

5 Results

5.1 Estimation of FMC with Sentinel-2 observations

To give an impression of the characteristics of the FMC estimation, the estimates from 6 days of Sentinel-2 observations throughout the year 2019 are visualized in Figure 11.

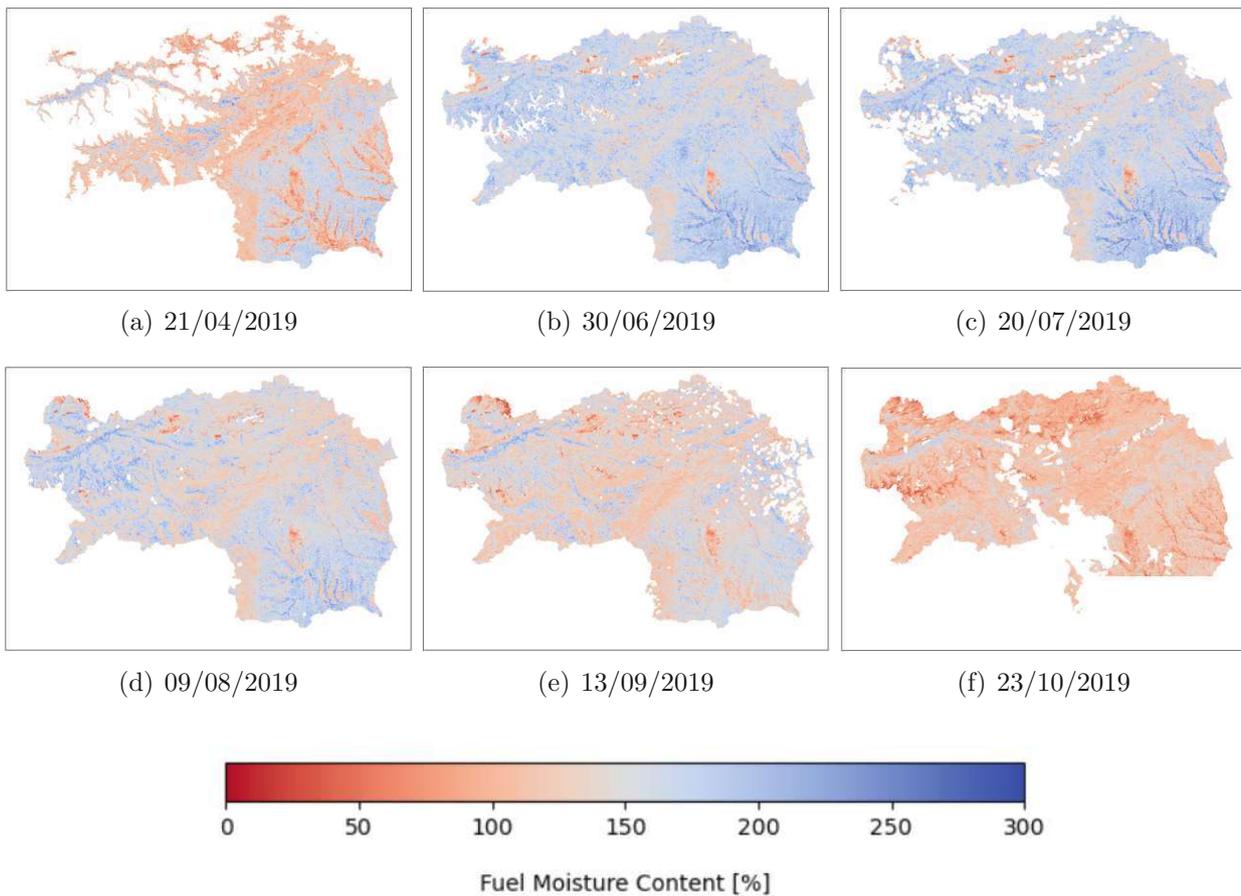


Figure 11: FMC estimations for Styria in 2019

It can be seen from the maps that there exist temporal dynamics in FMC over the Styrian territory. On 21/04/2019 (Figure 11a) the non-afforested valleys in the south of the research area seem to have the lowest FMC values, especially compared to their immediate surrounding area, which is dominated by forest. This is in contrast to the northern, alpine area, where the valleys in between the alpine ridges have higher FMC values than their forested surroundings with higher elevations. As summer starts, the FMC values are clearly increasing throughout the whole research area. The local differences between valleys and forests decrease in the northern area while in the south the valleys are now containing higher fuel moisture compared to the surrounding forest (see Figures 11b+c). With 09/08/2019 (Figure 11d), it can be seen that a drier period has started; only the valleys in the northwestern district of Liezen (see Figure 3) seem to save their fuel moisture. This dehydration goes on as fall begins. Especially the southern area dries out and the valleys there become the most arid regions in their surrounding area again, while the northern valleys remain as the areas with the highest FMC to that time (see Figures 11e+f).

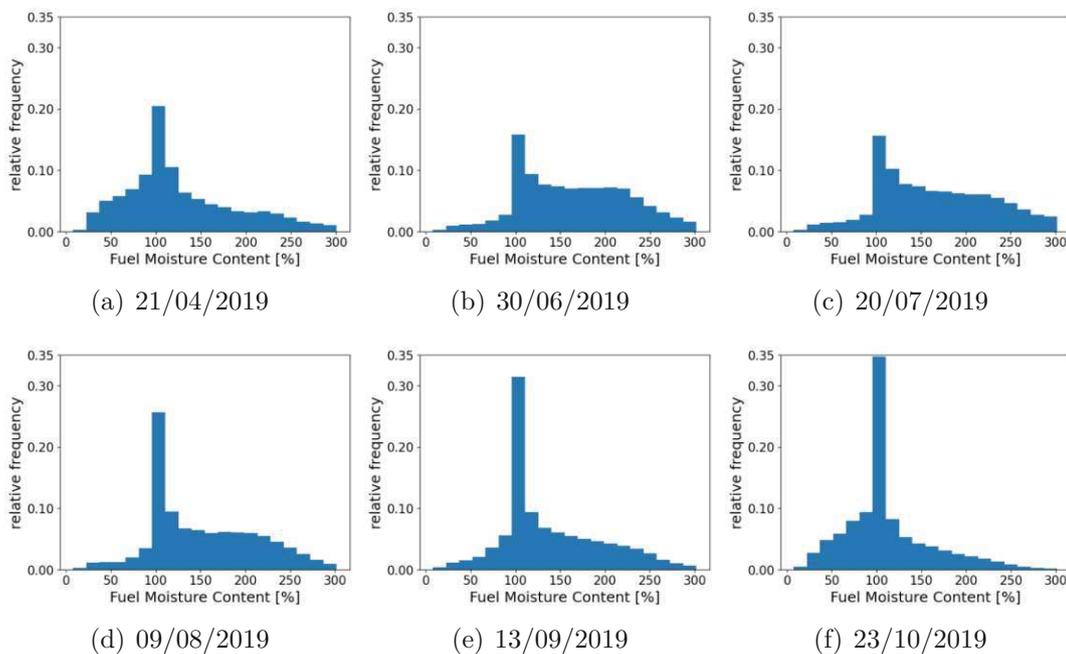


Figure 12: Distribution of Styrian FMC estimates

The dynamic processes in FMC throughout a year are also depicted in the histograms of the individual estimation dates (Figure 12). Remarkable is the fact that to every observation day, the maximum number of FMC estimations lies within the interval of 96 and 110%.

This spectrum of FMC is especially covered by densely forested areas (e.g. in the Kor alps (district Deutschlandsberg) or in the eastern district Hartberg-Fürstenfeld).

5.2 Area of applicability (AOA) of the random forest model

The calculation of the AOA of the random forest model (like described in Chapter 4.6) for three different Sentinel-2 observations in 2019 - each about three months apart - leads to the conclusion that for the major part of the Styrian territory, the model is applicable (Figure 13).

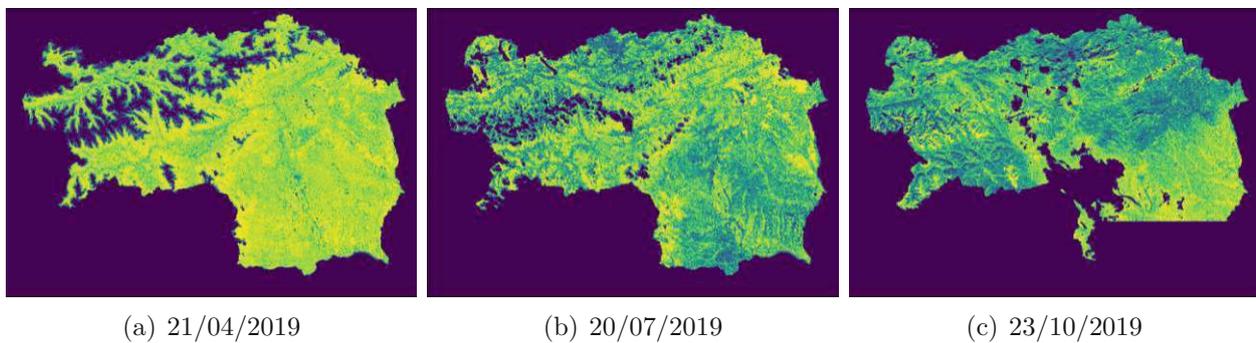


Figure 13: AOA of the random forest model for Styrian Sentinel-2 data - pixels coloured in yellow belong to the AOA

Especially in April, a huge majority of the non-masked data points is inside of the AOA, while this percentage is reduced for the samples in July and October. It is clarified that the Sentinel-2 image of the southern part of Styria on 23/10/2019 was omitted because of too much cloud cover. A more specific impression of the similarity between the Styrian data points and the Australian training data is given with Figure 14, where the calculated dissimilarity index values are plotted. Their maximum was limited to 200% of the threshold as there are a few outliers in the data. In April, cities and settled areas show the highest dissimilarity to the training data and therefore the lowest contribution to the AOA. Additionally, the northern valleys with higher FMC (see Figure 11a) contain relatively more pixels outside the AOA.

In July and October, the dissimilarity index shows more variation than in April. In addition to the settled areas, the northern part and also the southern part of Styria contain less similar data compared to the Australian training data in July. For October, the dissimilarity

in the north of Styria increases compared to July, but decreases in the southern part, where now a majority of the pixels is inside the AOA.

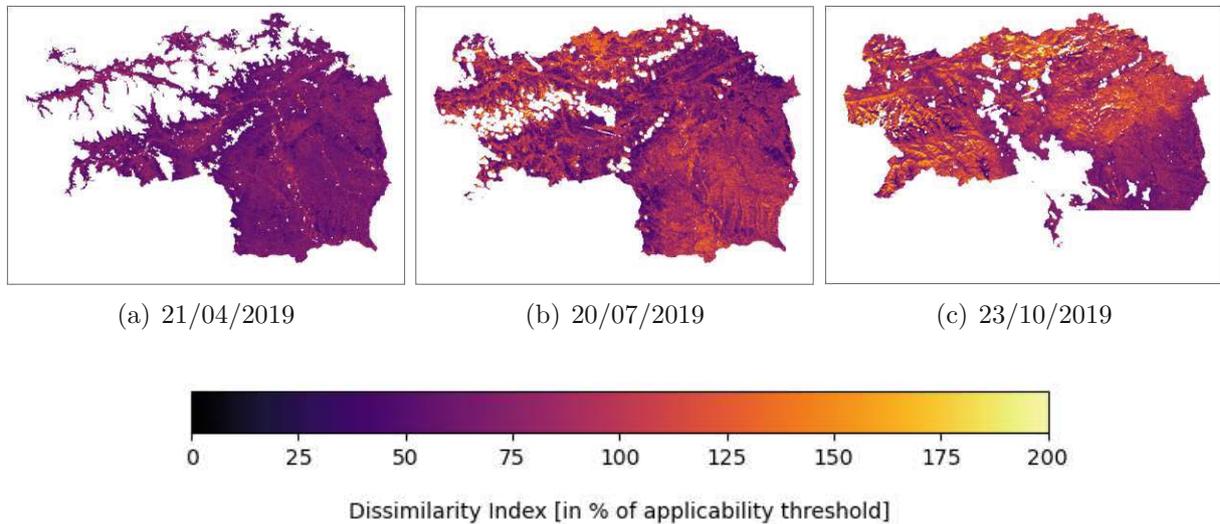


Figure 14: Map of the dissimilarity index for Styria

These conditions are also depicted in Figure 15, where the histograms of the dissimilarity index values are plotted. The more similar reflectance conditions in April stand out in comparison to the histograms of July and October, where significantly less data points are inside the AOA. Figure 15c has to be interpreted carefully, as the southern part of Styria is missing in the data processed.

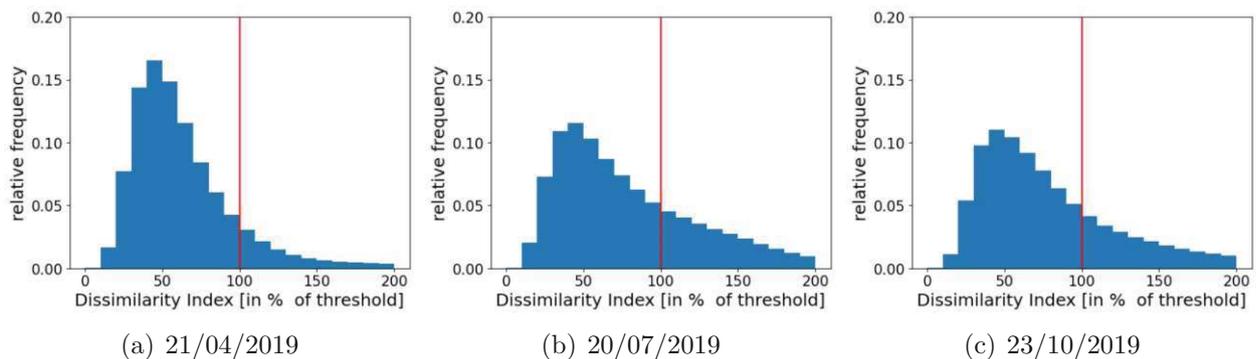


Figure 15: Histograms of the dissimilarity index values in % of the threshold - the threshold for belonging to the AOA is visualized in red

5.3 Correlation between Sentinel-2 FMC estimations and the MODIS FMC data product

5.3.1 Correlation between single observation days

In Figure 16, the compared datasets of Sentinel-2 based FMC estimations and the MODIS FMC data product are shown on top of each other. As an initial evaluation, it is remarkable that the MODIS FMC dataset has very high contrast in itself, basically the non-afforested valleys show very high FMC values up to 300% while the forested areas have similar FMC between about 100 and 150%. In comparison to the Sentinel-2 estimations, the MODIS dataset shows hardly any bigger temporal trend.

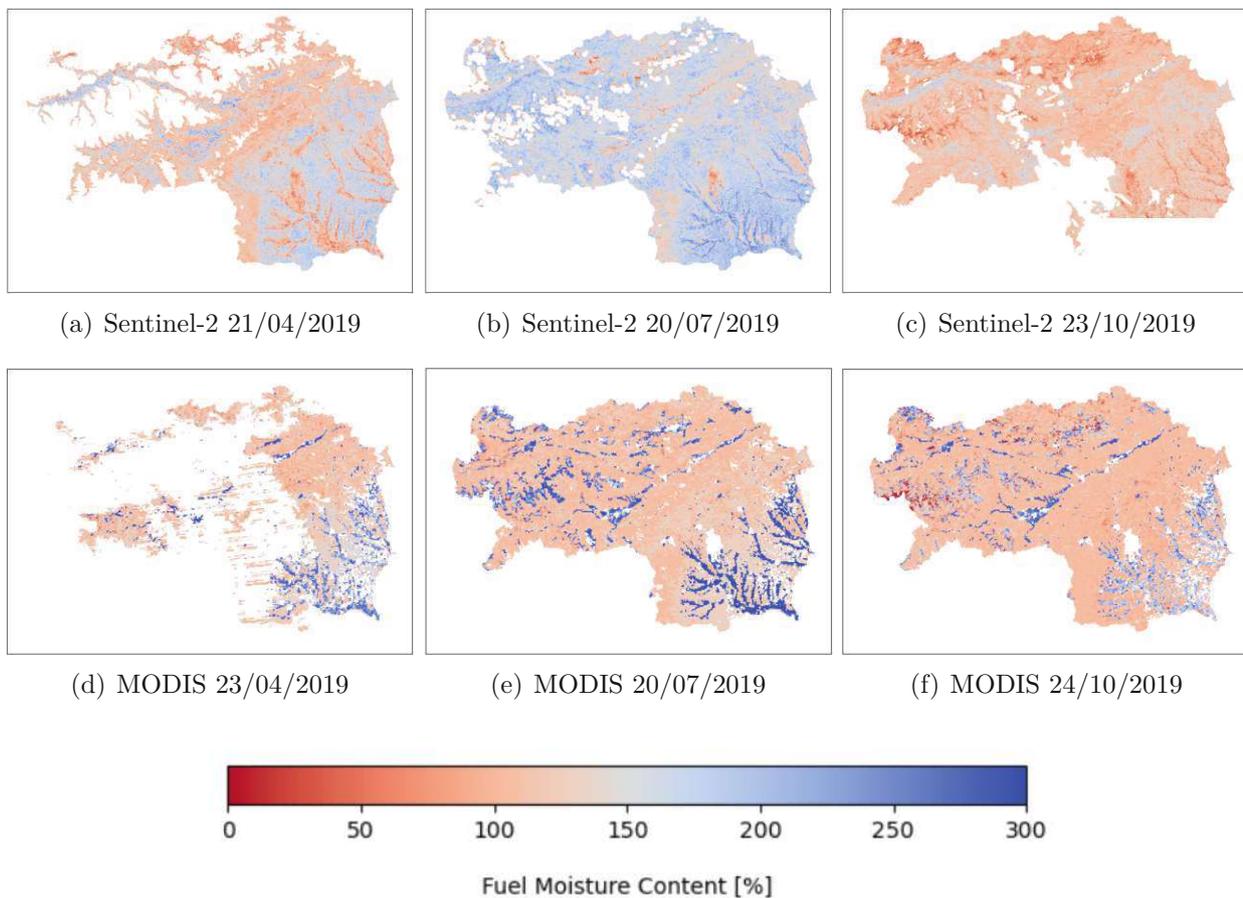
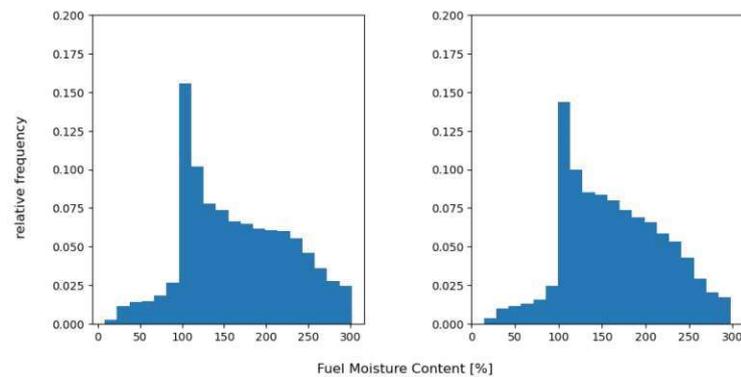


Figure 16: Compared FMC datasets from Sentinel-2 and MODIS

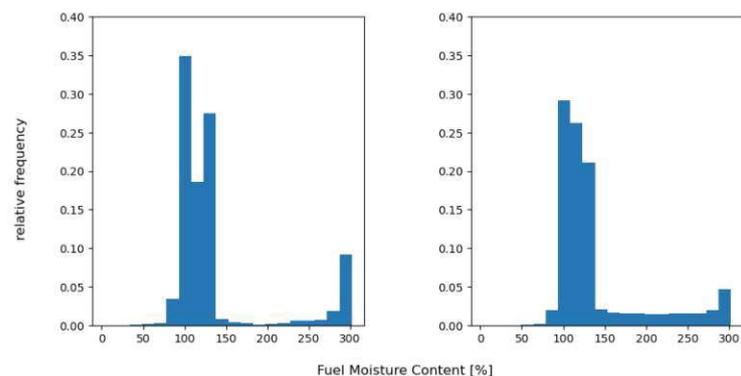
The major differences between those two FMC datasets that are affecting the correlation analysis are:

- The southern valleys have constantly very high FMC values for all three timestamps in the MODIS FMC dataset while this is only the case in July for the Sentinel-2 FMC estimations.
- The forested areas only show minor differences in FMC over the year for the MODIS dataset while a significant variation is clearly visible in the Sentinel-2 FMC estimations, especially between July and October.

Therefore, it can be already expected by visual evaluation that the correlation between those two FMC datasets is rather low.



(a) Sentinel-2 FMC estimations from 20/07/2019 before (left) and after (right) downsampling from 20 m to 500 m resolution



(b) MODIS FMC values from 20/07/2019 before (left) and after (right) upsampling from 500 m to 20 m resolution

Figure 17: Histograms of the FMC distribution before and after interpolation

Before calculating the correlation quantities, both interpolation methods of Chapter 4.7.1 were validated by checking the changes in distribution of FMC values (e.g. Figure 17). The shapes of the histograms for the Sentinel-2 FMC datasets are nearly identical. In contrast to that, the histogram of the MODIS FMC dataset shows a smoother trend after upsampling because more FMC values are interpolated into the range of about 140 and 280% between the local maxima, but the characteristics of the statistical distribution remain preserved for both sampling techniques.

The resulting correlation quantities between the two FMC datasets, calculated like described in Chapter 4.7.2, are listed in Table 1. It is visible that correlations are very low for all tested dates. The best result occurs for the datasets of 20/07/2019, where the southern valleys are the most humid areas in both datasets. The disregard of data points outside the AOA does not lead to a significant increase of the correlation coefficients, in two cases it is even lowering those quantities. Also the downsampling of the Sentinel-2 FMC estimations to 500 m - to reduce noise in the data - does not improve correlation considerably. As the computational effort for upsampling the MODIS data was lower due to less input data from the 500 m tiles than from a 20 m Sentinel-2 image, the high-resolution FMC datasets with 20 m pixel size were selected as the foundation for further analysis.

	20 m resolution				500 m resolution	
	without AOA mask		with AOA mask			
04/19	$\rho = 0.0613$	$R^2 = 0.0038$	$\rho = 0.0140$	$R^2 = 0.0002$	$\rho = 0.0597$	$R^2 = 0.0036$
07/19	$\rho = 0.1413$	$R^2 = 0.02$	$\rho = 0.1455$	$R^2 = 0.0212$	$\rho = 0.1591$	$R^2 = 0.0253$
10/19	$\rho = 0.0589$	$R^2 = 0.0035$	$\rho = 0.0184$	$R^2 = 0.0003$	$\rho = 0.0646$	$R^2 = 0.0042$

Table 1: Correlation quantities between the MODIS and the Sentinel-2 FMC dataset

The missing of a recognizable linear trend is also shown in the direct comparison of the two datasets via a 2D histogram (Figure 18). For the 20 m resolution, a minimum threshold of 500 data points per pixel was determined for the colourization, for 500 m resolution this value was reduced to 5. It is visible that the highest density of data points occurs in an equal range of FMC between 102 and 108% for both datasets. Nevertheless, as there are many data points where the two datasets differ significantly in their FMC estimation, this is not sufficient for getting good correlation results.

What also stands out in Figure 18b is the fact that the MODIS FMC data is distributed in a very unbalanced way with most of the values located between 90 and 140% and between 280 and 300% (like also shown in Figure 17b). The Sentinel-2 estimations are covering the range of FMC values more balanced without a local maximum on the upper border (see Figure 12).

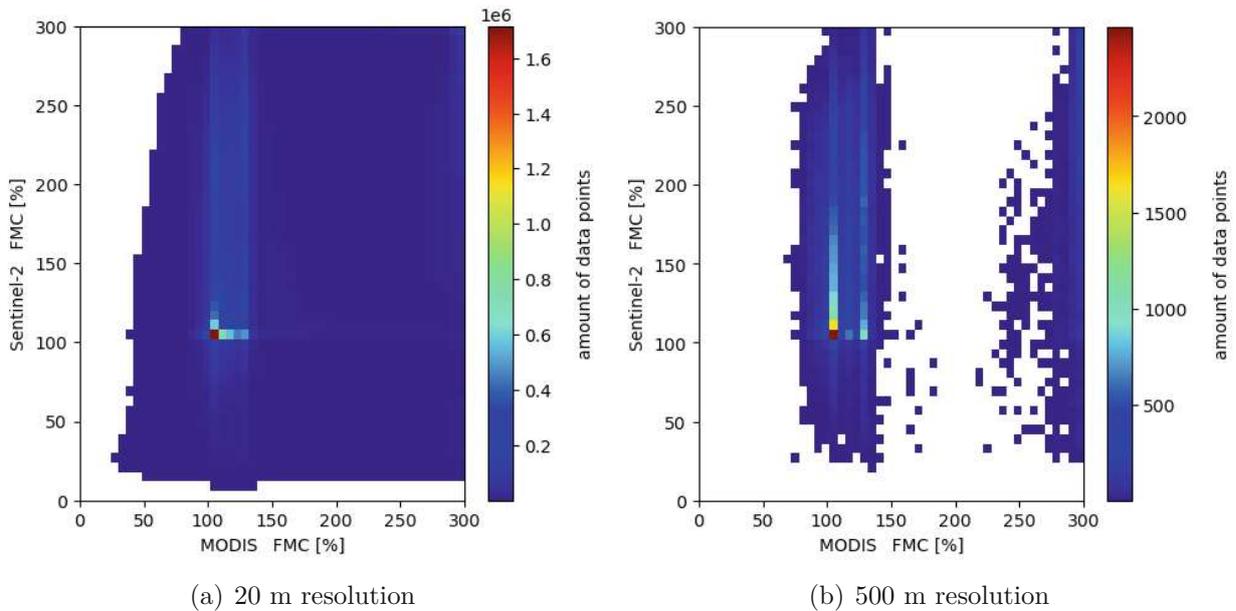


Figure 18: Colour-coded 2D histograms of the comparison between the MODIS FMC value and the associated FMC prediction from Sentinel-2 data on 20/07/2019

To clarify that this bad correlation is not caused by specific areas, where correlation is much lower than for other regions, the autocorrelation map for 20/07/2019 is shown exemplary in Figure 19. This map demonstrates that there are no significant local differences in correlation that may be caused by e.g. land cover type or vegetation. It shows a random pattern of areas with negative correlations to $\rho = -0.5$ and those with positive correlations up to 0.5 with the majority of values for ρ distributed closely around 0.

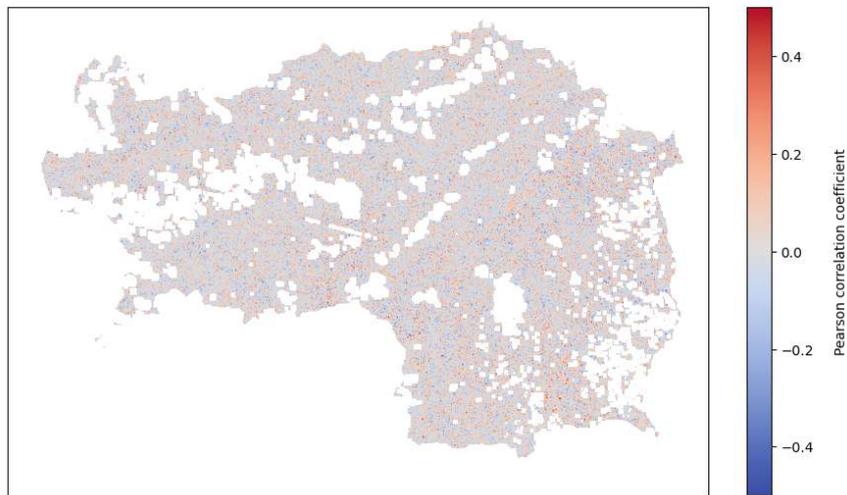


Figure 19: Local Autocorrelation map of Sentinel-2 and MODIS based FMC on 20/07/2019

5.3.2 Temporal correlation

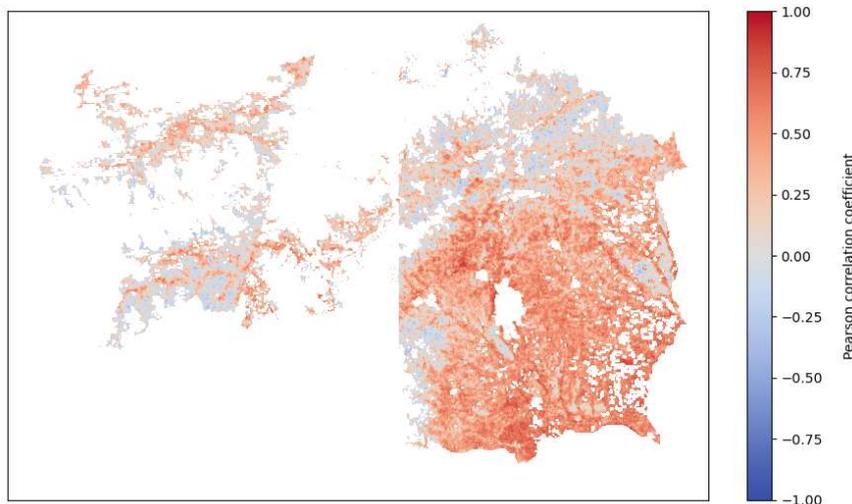


Figure 20: Temporal correlation of Sentinel-2 FMC estimations and MODIS FMC over 2019

The correlation of the FMC trends over 2019 of the Sentinel-2 and MODIS dataset is depicted in Figure 20. In contrast to the correlation between single observation days in Chapter 5.3.1, the positive correlations predominate in this analysis. Especially in the southern, less alpine part of Styria, correlations are rather high with a relevant amount of data points, for which ρ takes on values up to 0.75 (Figure 21). Negative correlations occur

especially in the densely forested area of the Kor alps (district Deutschlandsberg) and of the district Hartberg-Fürstenfeld, which have been already identified as areas with nearly constant Sentinel-2 FMC estimates between 96 and 110% in Chapter 5.1. In the northern part of Styria, it is visible that the correlations drop with increasing elevations. Only in the valleys with mostly non-forested areas, positive correlations are recognizable.

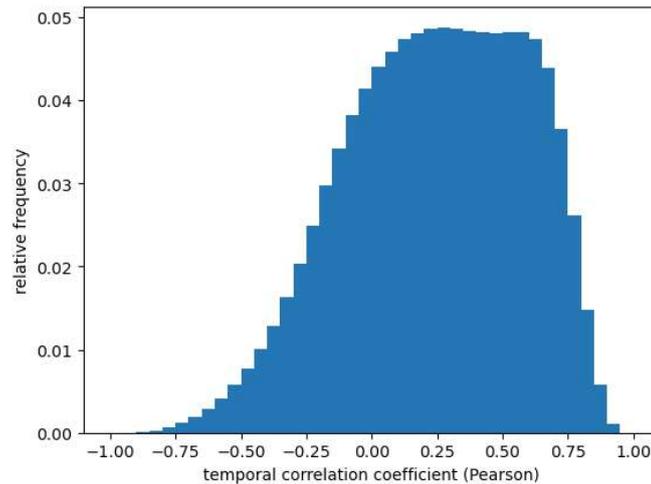


Figure 21: Histogram of the temporal correlation coefficients from Figure 20

5.4 Evaluation of the interrelation between wildfire events and FMC

In Figure 22, the classification results of the 50 Styrian wildfire events in 2019 are visualized based on their estimated FMC on the date of occurrence. For both datasets, it can be seen that only a little amount of the wildfires took place when the FMC value was located in the lowest 10% of the FMC values over 2019. For the MODIS dataset, at least 4 wildfires were classified to occur in the lowest 10%-quantile of FMC distribution while none of the wildfires were classified like that in the Sentinel-2 FMC estimation with the random forest model. 35 of the 50 wildfires were even classified to occur in the upper 50% of the Sentinel-2 FMC estimations while this was only the case for 22 wildfires in the comparison with MODIS FMC. Overall, it can be said that there is no correspondence of wildfire events with rather low FMC visible for both datasets. Especially for the Sentinel-2 FMC estimation, the wildfire events were classified to occur in the upper part of FMC values.

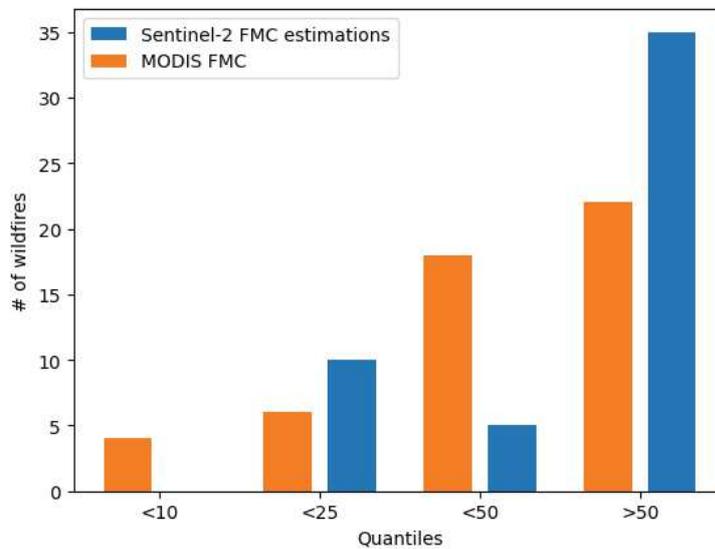
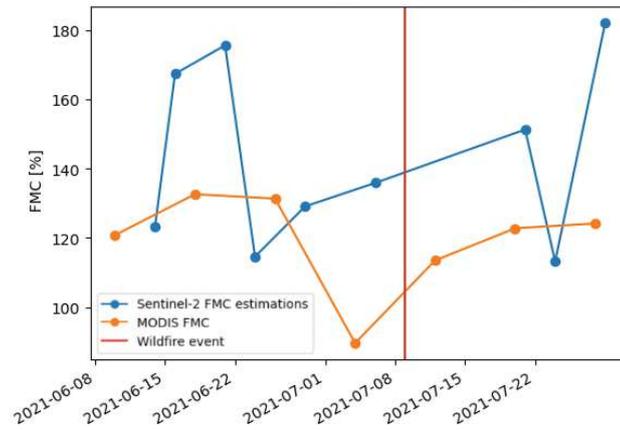
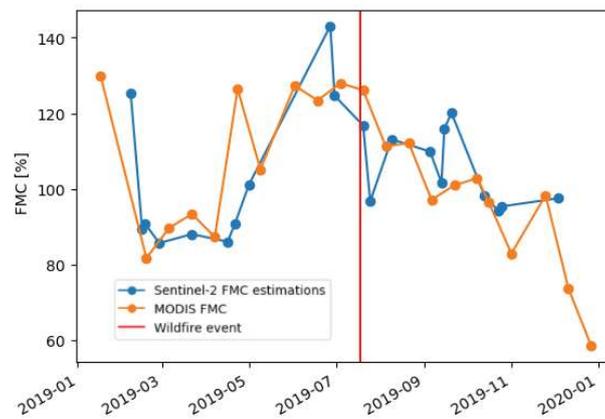


Figure 22: Occurrence of wildfires in Styria in 2019 related to FMC

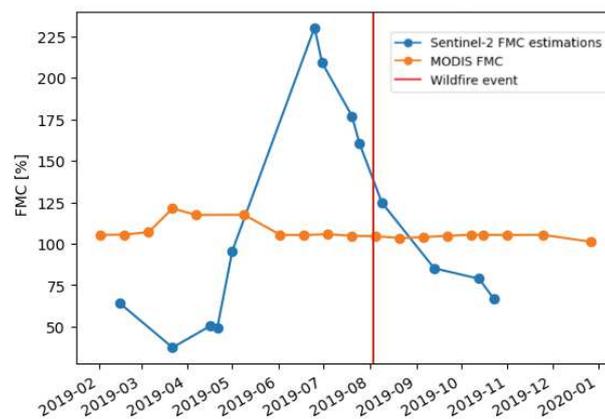
With having a closer insight on specific wildfire events like in Figure 23, different relations between the two FMC datasets among each other and together with the wildfire event can be seen. In case of the wildfire on 09/07/2021 (Fig. 23a), where a shorter period has been analyzed, the MODIS FMC dataset is detecting a decrease in FMC shortly before the event while the Sentinel-2 estimations show a different trend throughout the year and an increase in FMC before the wildfire. For the wildfire on 18/07/2019 (Fig. 23b), a similar trend with some outliers is shown for Sentinel-2 and MODIS based FMC values in the year of 2019. The wildfire occurs after a slight decrease in FMC in both datasets, but FMC was quite high in the annual course at this time with approximately 120%. The last example of a wildfire on 03/08/2019 (Fig. 23c) shows a bad correlation between the two FMC datasets. While the Sentinel-2 estimations lie in a wide range over the year and the wildfire occurs after a longer period of dehydration, the MODIS FMC data shows hardly any variation.



(a) Wildfire Grafendorf bei Hartberg (09/07/2021)



(b) Wildfire St. Radegund (18/07/2019)



(c) Wildfire Bruck an der Mur (03/08/2019)

Figure 23: FMC values of selected areas where wildfires occurred

6 Discussion

6.1 Deriving FMC from Sentinel-2 data

As the follow-on mission of Landsat, the use of reflectance data from the Sentinel-2 satellites to estimate FMC for wildfire monitoring is offering huge potential for current remote sensing techniques as it comes with a high spatial resolution of 20×20 m, which surpasses the resolutions of some established datasets like MODIS FMC. The temporal resolution of 2 to 3 days revisit time at mid-latitudes is also not much lower than those of the MODIS FMC dataset (500×500 m spatial resolution, 1-2 days revisit time). Furthermore, the comparison of the correlation results has shown that the high spatial resolution of the Sentinel-2 dataset had no negative effect in the validation compared to the downsampled dataset. Anyhow, as the validation of Sentinel-2 based FMC with the MODIS FMC dataset did not show significant positive correlations, this result should not be unquestioned for further research.

A huge limitation in this study has been the access of cloud-free Sentinel-2 data for Styria. The filtering threshold of 20% cloud cover for the satellite images lead to longer periods without any Sentinel-2 observations in this study (e.g. between 01/05/2019 and 12/06/2019), which would negatively affect the validity of wildfire danger assessment systems. Therefore, a compromise has to be found in the integration of Sentinel-2 based FMC estimations between high temporal frequency and high-quality cloud-free data, so that such long data gaps could be reduced without simple interpolation techniques as used in this thesis.

Another restriction for FMC estimation with Sentinel-2 data is a lack of simultaneous observations of Sentinel-2 reflectance and in-situ measurements of FMC. This is limiting the possibility to calibrate new models based on local environmental conditions. The harmonized Sentinel-2 dataset from Google Earth Engine is available from 28/03/2017 to the present [European Space Agency, 2023], while regular field measurements for multiple sites in Central Europe - where the environmental conditions are expected to be the closest to Styria - are

available in the Globe-LFMC database [Yebra et al., 2019] until 2016. Therefore, the option to train a new model for FMC estimation based on European field measurements could not be tested in this thesis. Recently, Yebra et al. [2024] have published an update of this database called Globe-LFMC 2.0, which would open up new opportunities for this field of research.

6.2 Applicability of the random forest model

An important factor of concern in the methodology of this thesis is the application of a model which was trained on Australian data to an area with arguably very different environmental and climatic conditions, as it is the case for Austria. In contrast to the climate of Styria (see Chapter 3), a large part of the territory of Australia is classified as arid (see Figure 24). This applies mainly for the interior of the country, which is categorized into the climates BWh (hot desert), BSh (hot semi-arid), BSk (cold semi-arid) and BWk (cold desert). However, there are specific regions along the coastlines where the classification is more comparable to Styria. At the eastern coast, where the density of training data points is rather high (see Figure 4), the climate is similar to the lower parts of Styria, as it is also fully humid with hot (Cfa) or warm summers (Cfb) [Kottek et al., 2006].

This is also the area with the highest proportion of forest in Australia according to its land cover map (Figure 25). Anyhow, the structure of the forests is highly different to those in Styria as more than 80% of the forest area in Australia is dominated by Eucalypt and Acacia forest [ABARES, 2015]. In addition to that, a large amount of training data points can be assigned to the open shrublands and savannas of Australia in the arid climate zone. Overall, this may not be a good reference for a machine learning model that is used on Styrian input data and is probably a major reason for the bad validation results.

Contrary to the climatic and vegetational differences, the calculation of the area of applicability based on a subset of the training data is leading to the impression that major parts of Styria are suitable for the use of the random forest model (see Chapter 5.2). However, it is visible that there are seasonal differences in the dissimilarity of the Styrian reflectance data to the Australian training data and therefore, the applicability of the model throughout a year is not guaranteed.

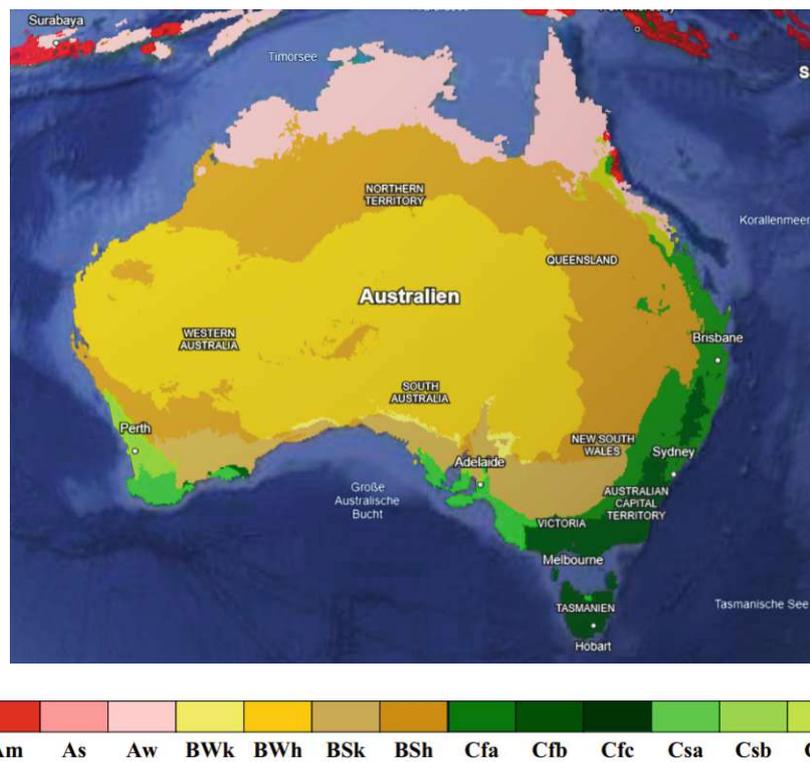


Figure 24: Köppen-Geiger Climate Classification of Australia [Kottek et al., 2006]

The methodology of Meyer and Pebesma [2021] has to be questioned in this use case since also pixels of urban areas are retaining the dissimilarity threshold, although the Australian training dataset only contains observations from the land cover types grassland, shrubland and forest. The consideration of the determined AOA in the correlation analysis is further not improving the validation results. The local differences in the dissimilarity that are creating a random pattern in Figure 13 could possibly have been reduced by averaging the 20×20 m reflectance data over a greater area like in the creation of the training dataset, but it is not likely that this would remove the mentioned irregularities.

The FMC estimations for Styria show some results that match the evaluations from Larraondo [2021]. For example, the high amount of FMC estimations between 96 and 110% in Figures 11 and 12 can be assigned to forested areas, while higher FMC was estimated for the valleys that are dominated by grass- or cropland. In the model training of Larraondo [2021] the FMC values up to 300% were also estimated for grassland, while for forest only a few data points exceeded a FMC estimation of 150%. Nevertheless, the suitability of the model for Styrian Sentinel-2 data has to be assessed critically because of the different

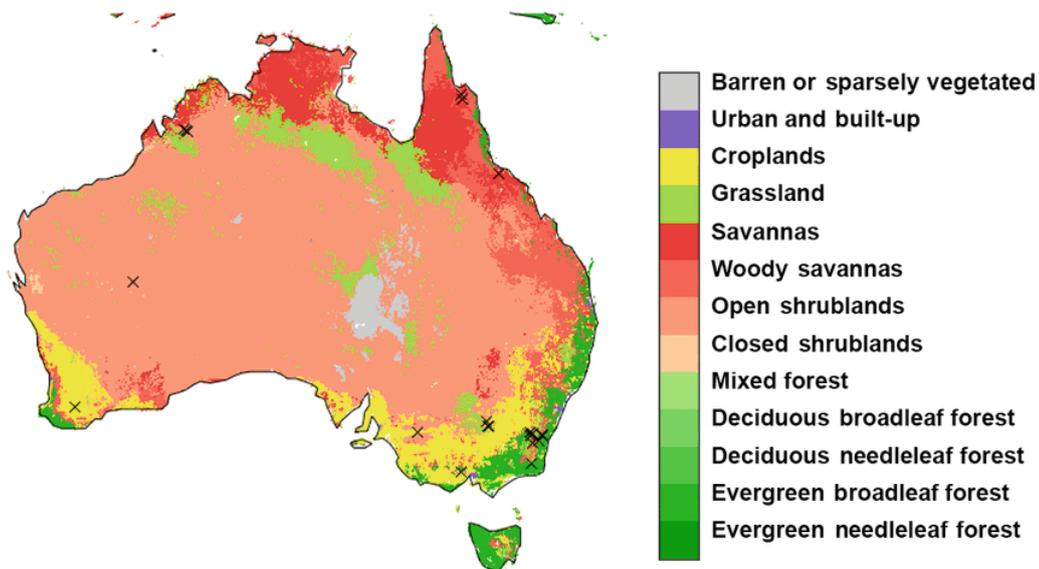


Figure 25: Land cover map of Australia [Chaivaranont et al., 2018]

bandwidth of FMC values in the training dataset. The major part of the training data is covering the range from 0 to 100% (see Figure 5), while the estimations for Styria are mainly over 100% (Figure 12). This condition may be the reason for some results in Chapter 5. The larger amount of FMC estimations below 100% on 21/04/2019 may explain the lower dissimilarity of the Styrian Sentinel-2 data to the training data compared to other observation dates (Figure 14). This is also underpinned by the fact that the valleys in Northern Styria with higher FMC contained more pixels outside the AOA on this date.

What has to be considered as well is the fact that the Sentinel-2 data from the datacube *Digital Earth Australia*, which has been used for training the random forest model, has passed a certain preprocessing workflow, which leads to small deviations from the Sentinel-2 reflectance data obtained from Google Earth Engine. This preprocessing algorithm is not known more detailed and has been neglected in this approach of FMC estimation for Styria, but may contribute to some extent to the not satisfying correlation results with MODIS FMC.

6.3 Validation with MODIS FMC data

The validation of Sentinel-2 FMC estimations with the MODIS FMC dataset has revealed very low correlations when single observation days have been compared. The conformity

of the maximum in the 2D-histogram (Figure 18) in the range from 102 to 108% of both datasets may not be caused by good correlation between the two datasets in this FMC range, but more likely by the fact that for both datasets the maximum of FMC values lies approximately in this range. This can be seen in Figure 17.

Figure 17 also shows the bad comparability of those two datasets caused by the different distribution of the respective FMC values. This harsh contrast between high FMC (up to 300%) in valleys dominated by grass- and cropland and much lower ones (around 90 to 140%) in the other areas is depicted through the entire year in the MODIS FMC dataset (see Figure 16). There is hardly any seasonal effect visible in the MODIS datasets, while the variation of the Sentinel-2 based FMC estimations throughout the year 2019 is significantly higher, which opens up the question: which trend is more realistic? This could be answered with frequent in-situ FMC measurements of the same samples in Styria to investigate the behaviour of the vegetation with regard to water storage and dehydration over the yearly cycle.

The unbalanced distribution of MODIS FMC data is also seen in the training sample of Australia (Figure 5), which is based on the MODIS data product. As the distribution varies between the Styrian and Australian FMC data especially in the range from 0 to 100%, it is likely that an empirical fitting of a machine learning algorithm can not work for both countries. The fact that the MODIS FMC data has been filtered for sites with good data quality in the training dataset may be a reason for worse correlations when applying the model to a greater area. The circumstance that the MODIS FMC dataset is based on radiative transfer models of grassland, shrubland and forest [Quan et al., 2021b] is also leaving the question, why the FMC of vegetational land covers like e.g. cropland should be estimated with this approach. These areas (concretely the valleys in southern Styria - see Figure 3) are not filtered in the MODIS FMC data, but are showing the inverse trend in comparison to the Sentinel-2 based FMC estimation (see Figure 16).

However, this idea is not supported by the local autocorrelation map (Figure 19), as it states that there are no smaller areas with potentially identical land cover which show a higher positive correlation. A comparable validation result of Australian Sentinel-2 estimations from the random forest model to the MODIS FMC dataset would be of interest, if the model can be also well validated for the application on greater areas and not only for the test sample from the optimized training data. These ideal conditions of the training dataset,

which is among other things based on in-situ measurements of Australia [Quan et al., 2021b] and thus the local climatic and vegetational conditions, together with the site-specificity of empirical machine learning models (as discussed in Chapter 2.2.3) is expected to be a major contributor to the bad validation results of the derived Sentinel-2 FMC values in Styria with MODIS FMC data.

The analysis of temporal correlation between the Sentinel-2 FMC estimations and the MODIS FMC dataset is showing more promising results. Apparently, the unequal spacing of the Sentinel-2 datasets, caused by the cloud cover filtering as mentioned in Chapter 6.1, is not an ideal condition for comparing the two datasets as interpolation can be based on estimations further away in terms of time. Nevertheless, the amount of visible positive correlations in the research area is significantly higher than for the single day analysis and therefore changes in FMC are detected similarly by both FMC derivation methods for many regions.

When comparing the temporal correlation map (Figure 20) with the land cover map of Styria (Figure 3), it is visible that correlations are dependent of the land cover type. For valleys dominated by grassland but also by cropland, correlations are higher, which corresponds with the result of Larraondo [2021] when testing the random forest model. Higher correlations ($R^2 = 0.7$) were achieved for grassland there than for forest ($R^2 = 0.47$) or shrubland ($R^2 = 0.41$). Densely forested areas, that are also clearly visible in satellite imagery (Figure 7), are determined to be vegetation types where the random forest model is definitely not applicable as temporal correlations are negative for this land cover type. This is probably also caused by the different tree species in both countries, as mentioned in Chapter 6.2.

6.4 Validation with wildfire data

Before analyzing the occurrence of wildfires in relation to FMC, it was assumed that wildfires are more likely to appear when FMC is rather low. Figure 22 shows that both FMC datasets are not supporting this assumption. The relative majority of wildfire events was assigned to happen in the upper 50% of local FMC values in 2019.

The initial assumption may not hold as the most important reasons of wildfire ignitions, as lightnings and anthropological causes, are expected to appear in a random pattern of time and location, which is not possible to predict. Müller and Vacik [2017] have even

stated the opposite conclusion, that an increase in fuel moisture - based on the Fine Fuel Moisture Code of the Canadian Fire Weather Index - was detected when analyzing FMC in the period from about 10 days before lightning-caused wildfire events.

The flammability index of Yebra et al. [2018] was based on a logistic regression model trained with pairs of MODIS FMC values of burned areas and spatially close unburned pixels right before a wildfire event. The predictor variables for model training were built up with the FMC values of the two 8-day data periods before burning, the difference of the FMC values of those two periods and a parameter for the anomaly of those FMC values compared to the time series from 2001 to 2016. This approach was also designed with the expectation of wildfires occurring when FMC decreases and achieved good results with "Area under the Curve" values from the ROC (receiver operating characteristic) plot between 0.7 and 0.8 for separate subsets.

So, the statement that wildfires are occurring more likely in periods with lower and decreasing FMC has been as well supported and weakened in previous research work. The individual wildfires of 2019 in Styria, with some examples shown in Figure 23, do not reveal a systematic pattern to evaluate this assumption. A more frequent data availability of cloud-free Sentinel-2 observations would enable to perform a similar analysis like Yebra et al. [2018] in prospective research work and compare related quantities of wildfire prediction.

6.5 Potential improvements

Following on from the previous chapters, there is some room for improvement left that should be considered for future applications or subsequent studies. In the collection of Sentinel-2 data for further processing, the possibility of masking cloudy areas in the satellite images pixel-wise instead of using an overall threshold would increase the amount of available data and reduce long gaps in the input data of the random forest model. Additionally, a land cover map should be integrated in the pixel masking to estimate FMC only for areas covered by vegetation. Figure 11 reveals that constantly low FMC values are estimated for mountainous areas above the vegetation line, as well as for urban areas and lakes. These pixels could have also been masked as they do not contribute to wildfire ignition and propagation. In this study, this has been neglected as in the validation with the MODIS FMC dataset, non-vegetational land cover types have been masked anyway.

The FMC estimation for more Styrian Sentinel-2 datasets with reduced time intervals would also improve the reliability of the temporal correlation analysis with MODIS FMC. Furthermore, a separation of the datasets with regard to the land cover types grassland, shrubland, forest and - as a new component - cropland, can give more insight into the correlation between the two datasets and the suitability of the random forest model for different vegetation types.

When working with the random forest model of Larraondo [2021], the preprocessing workflow of *Digital Earth Australia* should be considered and applied to the extracted Sentinel-2 data from Google Earth Engine. To assess the applicability of this model more critically, a different approach for the evaluation of the applicability of a model should be reviewed. A combined analysis of Sentinel-2 data with other sensors, which are operating in the microwave spectrum (e.g. Sentinel-1), may give a better representation of the climatic and vegetational differences between two areas in order to decide if a model is applicable to a specific region or not.

Due to the climatic and vegetational differences between Australia and Styria, a valid workflow for Sentinel-2 based FMC estimation for Austria probably can be only expected, when a new model is trained on Austrian conditions or at least on similar ones. This could be done with in-situ FMC measurements from Central Europe. The database *Globe-LFMC* recently has been updated, therefore, new spatially well distributed and temporally frequent observations exist, that are overlapping in time with available Sentinel-2 data from Google Earth Engine [Yebra et al., 2024]. Alternatively, optical reflectance data from a different satellite could be chosen for developing a new FMC model.

As long as such a model does not exist, the prototype of an Austrian wildfire danger assessment system from Müller et al. [2020b] possibly could be improved by substituting the Fine Fuel Moisture Code from the Canadian Fire Weather Index with the MODIS FMC dataset from Quan et al. [2021b]. This approach may be promising as the MODIS FMC is based on direct observations of water content in vegetation and not on meteorological models that were developed for a specific tree species in Canada. An evaluation of the change in the predictive capabilities of this fire danger assessment system may lead to more accurate forecasts of wildfire ignitions and their rate of spread in Austria.

7 Conclusion and outlook

The aim of estimating well-validated FMC values from Sentinel-2 observations was not fully achieved in this thesis. The properties of the training dataset and workflow from the random forest model of Larraondo [2021] were not sufficiently transferable to Styrian Sentinel-2 input data. The main reasons for the low correlation results when comparing Sentinel-2 FMC estimations based on the random forest model to the MODIS FMC dataset were identified as the following: (i) climatic and vegetational differences between Australia and Styria and the resulting site-specificity of the random forest model, (ii) different Sentinel-2 reflectance input data for the model training and the Styrian application due to the preprocessing algorithm of *Digital Earth Australia*.

For future approaches in FMC estimation for Austria, it should be considered to train a new machine learning model with in-situ measurements from selected sites with more comparable climatic conditions. In the Globe-LFMC database, the amount of measurements in Central Europe from 2017 on is not sufficient to compare this data to Sentinel-2 reflectances. Hence, observations from other multispectral data sources like Landsat-8 are needed to increase the timely overlap between the field measurements and remote sensing data. As a new data source, the lately published database Globe-LFMC 2.0 from Yebra et al. [2024] includes more recent in-situ FMC measurements, which opens up the possibility to train new models that are estimating FMC based on Sentinel-2 data and in-situ measurements from Central Europe.

To achieve improvements for wildfire danger assessment in Austria for now, the approach to integrate the MODIS FMC dataset from Quan et al. [2021b] in the prototypical fire danger assessment system of Müller et al. [2020b] - instead of the Fine Fuel Moisture Code from the Canadian Fire Weather Index - could be tested. In this way, daily updated FMC values derived from satellite observations can be considered alternatively to an index based on meteorological models calibrated for Canadian conditions.

A Observation days of used Sentinel-2 and MODIS data

Sentinel-2			MODIS
date	covered area	tiles filtered	date
11/01/2019	all	western tile	01/01/2019
07/02/2019	eastern and western strip		17/01/2019
15/02/2019	all		02/02/2019
17/02/2019	eastern and western strip		18/02/2019
27/02/2019	eastern and western strip		06/03/2019
17/03/2019	all	northeastern tile	22/03/2019
22/03/2019	all		07/04/2019
30/03/2019	all	central tile	23/04/2019
16/04/2019	all	western tile	09/05/2019
21/04/2019	all		25/05/2019
01/05/2019	all		02/06/2019
12/06/2019	eastern and western strip		18/06/2019
25/06/2019	all	southern tile	04/07/2019
27/06/2019	eastern and western strip		20/07/2019
30/06/2019	all		05/08/2019
20/07/2019	all		21/08/2019
25/07/2019	all		06/09/2019
09/08/2019	all		22/09/2019
05/09/2019	eastern strip		08/10/2019
13/09/2019	all		16/10/2019
15/09/2019	eastern and western strip		01/11/2019
20/09/2019	eastern and western strip		25/11/2019

APPENDIX A. OBSERVATION DAYS OF USED SENTINEL-2 AND MODIS DATA

13/10/2019	all	western tile	11/12/2019
23/10/2019	all	southern tile	27/12/2019
25/10/2019	eastern and western strip		
04/12/2019	eastern and western strip	southern tile	
14/06/2021	all		
16/06/2021	eastern strip		
21/06/2021	eastern strip		
24/06/2021	all		
29/06/2021	all	central tile	
06/07/2021	eastern strip		
21/07/2021	eastern strip		
24/07/2021	all		
29/07/2021	all		

Table 2: dates of Sentinel-2 observation days and MODIS FMC datasets used for FMC estimation and validation

List of Figures

1	Construction process of the random forest, built up by classification and regression trees (CARTs) [Sun et al., 2024]	9
2	Included data layers in the prototype of the integrated fire danger assessment system [Müller et al., 2020b]	10
3	Land cover map of Styria [Schulatlas Steiermark, 2018]	12
4	FMC data points used for training the random forest model	15
5	Distribution of FMC values in the training data	16
6	feature importance for estimating FMC with the Australian random forest model	17
7	RGB composite of a filtered Sentinel-2 image of Styria	18
8	Estimation of FMC (right) from a Sentinel-2 image (left as RGB composite) of 29/07/2021	22
9	Upsampling of a original 500 m MODIS FMC dataset (left) to the Sentinel-2 grid of 20 m resolution (right)	25
10	Downsampling of the Sentinel-2 FMC estimations with 20 m pixel size(left) to the MODIS FMC grid of 500 m resolution (right)	26
11	FMC estimations for Styria in 2019	30
12	Distribution of Styrian FMC estimates	31
13	AOA of the random forest model for Styrian Sentinel-2 data - pixels coloured in yellow belong to the AOA	32
14	Map of the dissimilarity index for Styria	33
15	Histograms of the dissimilarity index values in % of the threshold - the threshold for belonging to the AOA is visualized in red	33
16	Compared FMC datasets from Sentinel-2 and MODIS	34
17	Histograms of the FMC distribution before and after interpolation	35

List of Figures

18	Colour-coded 2D histograms of the comparison between the MODIS FMC value and the associated FMC prediction from Sentinel-2 data on 20/07/2019	37
19	Local Autocorrelation map of Sentinel-2 and MODIS based FMC on 20/07/2019	38
20	Temporal correlation of Sentinel-2 FMC estimations and MODIS FMC over 2019	38
21	Histogram of the temporal correlation coefficients from Figure 20	39
22	Occurrence of wildfires in Styria in 2019 related to FMC	40
23	FMC values of selected areas where wildfires occurred	41
24	Köppen-Geiger Climate Classification of Australia [Kottek et al., 2006] . . .	44
25	Land cover map of Australia [Chaivaranont et al., 2018]	45

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Tables

1	Correlation quantities between the MODIS and the Sentinel-2 FMC dataset	36
2	dates of Sentinel-2 observation days and MODIS FMC datasets used for FMC estimation and validation	52

Bibliography

- ABARES (2015). Australia's forests at a glance 2015. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.
- Benesty, J., Chen, J., and Huang, Y. (2008). On the Importance of the Pearson Correlation Coefficient in Noise Reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(4):757–765.
- Boston, T., Van Dijk, A., Larraondo, P. R., and Thackway, R. (2022). Comparing CNNs and Random Forests for Landsat Image Segmentation Trained on a Large Proxy Land Cover Dataset. *Remote Sensing*, 14(14).
- Bowyer, P. and Danson, F. (2004). Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. *Remote Sensing of Environment*, 92(3):297–308. Forest Fire Prevention and Assessment.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.
- Breitfelder, J., Pirker, E., Lieb, M., Ebner, B., Krobath, M., Pirker, D., Proyer, H., Pink, R., Sulzer, W., and Lieb, C. (2019). *Regionalgeographisches Kurzportrait der Steiermark*. Schulatlas Steiermark.
- Byram, George M.; Jemison, G. M. (1943). Solar radiation and forest fuel moisture. *Journal of Agricultural Research*, 67(4).
- Chaivaranont, W., Evans, J., Liu, Y., and Sharples, J. (2018). Estimating grassland curing with remotely sensed data. *Natural Hazards and Earth System Sciences*, 18:1535–1554.
- Cucca, B., Recanatesi, F., and Ripa, M. (2020). Evaluating the Potential of Vegetation Indices in Detecting Drought Impact Using Remote Sensing Data in a Mediterranean Pinewood. *Computational Science and Its Applications – ICCSA 2020*, 12253:50–62.

- De Angelis, A., Ricotta, C., Conedera, M., and Pezzatti, G. (2015). Modelling the Meteorological Forest Fire Niche in Heterogeneous Pyrologic Conditions. *PLoS ONE* 10(2).
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120:25–36. The Sentinel Missions - New Opportunities for Science.
- Dzugan, F. (2022). Waldbrand in Hirschwang an der Rax: Was wächst nach der Feuersbrunst? Profil. URL: <https://www.profil.at/wissenschaft/waldbrand-in-hirschwang-an-der-rax-was-waechst-nach-der-feuersbrunst/402181767> (last visit: 27/12/2023).
- European Space Agency (2015). Second Copernicus environmental satellite safely in orbit. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Second_Copernicus_environmental_satellite_safely_in_orbit (last visit: 05/08/2023).
- European Space Agency (2016). Sentinel-2 global coverage. ESA/ATG medialab. URL: https://www.esa.int/ESA_Multimedia/Videos/2016/08/Sentinel-2_global_coverage (last visit: 24/12/2023).
- European Space Agency (2023). Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. URL: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED (last visit: 20/02/2024).
- Forkel, M., Schmidt, L., Zotta, R., Dorigo, W., and Yebra, M. (2022). Estimating leaf moisture content at global scale from passive microwave satellite observations of vegetation optical depth. *Hydrology and Earth System Sciences*, 27:39–68.
- Fourty, T. and Baret, F. (1997). Vegetation water and dry matter contents estimated from top-of-the-atmosphere reflectance data: A simulation study. *Remote Sensing of Environment*, 61(1):34–45.
- García, M., Riaño, D., Yebra, M., Salas, J., Cardil, A., Monedero, S., Ramirez, J., Martín, M., Vilar, L., Gajardo, J., and Ustin, S. (2020). A Live Fuel Moisture Content Product from Landsat TM Satellite Time Series for Implementation in Fire Behavior Models. *Remote Sensing*, 12:1714.

- Gillies, S. et al. (2013). Rasterio: geospatial raster I/O for Python programmers. Mapbox. URL: <https://github.com/rasterio/rasterio> (last visit: 08/12/2023).
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- Kalabokidis, K., Athanasis, N., Gagliardi, F., Karayiannis, F., Palaiologou, P., Parastatidis, S., and Vasilakos, C. (2013). Virtual Fire: A web-based GIS platform for forest fire control. *Ecological Informatics*, 16:62–69.
- Kondylatos, S., Prapas, I., Camps-Valls, G., and Papoutsis, I. (2023). Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean. arXiv preprint arXiv:2306.05144.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15:259–263.
- Laerd Statistics (2020). Pearson’s product moment correlation. Statistical tutorials and software guides. URL: <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php> (last visit: 09/12/2023).
- Land Processes Distributed Active Archive Center (2023). MCD43A4 v006 - MODIS/Terra+Aqua Nadir BRDF-Adjusted Reflectance (NBAR) Daily L3 Global 500 m SIN Grid. URL: <https://lpdaac.usgs.gov/products/mcd43a4v006/> (last visit: 23/11/2023).
- Larraondo, P. R. (2021). Sentinel2 FMC workflow. GitHub. URL: https://github.com/ANU-WALD/sentinel2_fmc (last visit: 13/12/2023).

- McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.
- Meyer, H. and Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12:1620–1633.
- Müller, M. M. and Vacik, H. (2017). Characteristics of lightnings igniting forest fires in Austria. *Agricultural and Forest Meteorology*, 240-241:26–34.
- Müller, M. M., Vacik, H., and Valsecchi, E. (2015). Anomalies of the Austrian Forest Fire Regime in Comparison with Other Alpine Countries: A Research Note. *Forests*, 6(4):903–913.
- Müller, M. M., Vilà-Vilardell, L., and Vacik, H. (2020a). Forest fires in the Alps – State of knowledge, future challenges and options for an integrated fire management. *EUSALP Action Group 8*.
- Müller, M. M., Vilà-Vilardell, L., and Vacik, H. (2020b). Towards an integrated forest fire danger assessment system for the European Alps. *Ecological Informatics*, 60:101151.
- National Aeronautics and Space Administration (2023). MODIS Web. URL: <https://modis.gsfc.nasa.gov/> (last visit: 23/11/2023).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Quan, X., Li, Y., He, B., Cary, G. J., and Lai, G. (2021a). Application of Landsat ETM+ and OLI Data for Foliage Fuel Load Monitoring Using Radiative Transfer Model and Machine Learning Method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5100–5110.
- Quan, X., Yebra, M., Riaño, D., He, B., Lai, G., and Liu, X. (2021b). Global fuel moisture content mapping from MODIS. *International Journal of Applied Earth Observation and Geoinformation*, 101.

- Rao, K., Williams, A. P., Diffenbaugh, N. S., Yebra, M., Bryant, C., and Konings, A. G. (2023). Dry Live Fuels Increase the Likelihood of Lightning-Caused Fires. *Geophysical Research Letters*, 50(15).
- Riano, D., Vaughan, P., Chuvieco, E., Zarco-Tejada, P., and Ustin, S. (2005). Estimation of fuel moisture content by inversion of radiative transfer models to simulate equivalent water thickness and dry matter content: analysis at leaf and canopy level. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):819–826.
- Rossa, C., Veloso, R., and Fernandes, P. (2016). A laboratory-based quantification of the effect of live fuel moisture content on fire-spread rate. *International Journal of Wildland Fire*, 25:569–573.
- SchulAtlas Steiermark (2018). URL: www.schulatlas.at (last visit: 08/03/2024).
- Snow, A. D. et al. (2021). pyproj4/pyproj: 3.2rc0. URL: <https://zenodo.org/records/5348522> (last visit: 08/12/2023).
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., and Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237:121549.
- The pandas development team (2020). pandas-dev/pandas: Pandas. Zenodo. URL: <https://doi.org/10.5281/zenodo.3509134> (last visit: 10/12/2023).
- Tiede, D., Sudmanns, M., Augustin, H., and Baraldi, A. (2021). Investigating ESA Sentinel-2 products' systematic cloud cover overestimation in very high altitude areas. *Remote Sensing of Environment*, 252:112163.
- Vacik, H. (2023). Forest Fire Database Austria. Institute of Silviculture. University of Natural Resources and Life Sciences Vienna. Department of Forest and Soil Sciences. URL: <https://fire.boku.ac.at/firedb/en/> (last visit: 23/12/2023).
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Van Wagner, C. (1972). *Equilibrium moisture contents of some fine forest fuels in eastern Canada*. Canadian Forestry Service, Petawawa Forest Experiment Station, Chalk River, Ontario.

- Van Wagner, C. et al. (1987). *Development and structure of the Canadian forest fire weather index system*, volume 35. National Capital Region.
- Viney, N. R. (1992). *Modelling fine fuel moisture content*. PhD thesis, UNSW Sydney.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Wang, L., Quan, X., He, B., Yebra, M., Xing, M., and Liu, X. (2019). Assessment of the Dual Polarimetric Sentinel-1A Data for Forest Fuel Moisture Content Estimation. *Remote Sensing*, 11(13).
- Whitaker, J. et al. (2019). Unidata/netcdf4-python: Version 1.5.3 release. URL: <https://zenodo.org/records/3516272> (last visit: 08/12/2023).
- Yebra, M., Dennison, P. E., Chuvieco, E., Riaño, D., Zylstra, P., Hunt, E. R., Danson, F. M., Qi, Y., and Jurdao, S. (2013). A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. *Remote Sensing of Environment*, 136:455–468.
- Yebra, M., Quan, X., Riaño, D., Rozas Larraondo, P., van Dijk, A. I., and Cary, G. J. (2018). A fuel moisture content and flammability monitoring methodology for continental Australia based on optical remote sensing. *Remote Sensing of Environment*, 212:260–272.
- Yebra, M., Scortechini, G., Adeline, K., Aktepe, N., Almoustafa, T., Bar-Massada, A., Beget, M. E., Boer, M., Bradstock, R., Brown, T., Castro, F. X., Chen, R., Chuvieco, E., Danson, M., Degirmenci, C. U., Delgado-Dávila, R., Dennison, P., Di Bella, C., Domenech, O., Féret, J.-B., Forsyth, G., Gabriel, E., Gagkas, Z., Gharbi, F., Granda, E., Griebel, A., He, B., Jolly, M., Kotzur, I., Kraaij, T., Kristina, A., Kütük, P., Limousin, J.-M., Martín, M. P., Monteiro, A. T., Morais, M., Moreira, B., Mouillot, F., Msweli, S., Nolan, R. H., Pellizzaro, G., Qi, Y., Quan, X., Resco de Dios, V., Roberts, D., Tavsanoğlu, C., Taylor, A. F. S., Taylor, J., Tufekcioglu, I., Ventura, A., and Younes Cardenas, N. (2024). Globe-LFMC 2.0, an enhanced and updated dataset for live fuel moisture content research. *Sci Data*, 11:332.

Yebera, M., Scortechini, G., Badi, A., Beget, M. E., Boer, M., Bradstock, R., Chuvieco, E., Danson, F., Dennison, P., Resco de Dios, V., Di Bella, C., Forsyth, G., Frost, P., García, M., Hamdi, A., He, B., Jolly, W., Kraaij, T., Martín, M., and Ustin, S. (2019). Globe-LFMC, a global plant water status database for vegetation ecophysiology and wildfire applications. *Scientific Data*, 6:e155.

Zhang, P., Huang, Y., Shekhar, S., and Kumar, V. (2003). Correlation Analysis of Spatial Time Series Datasets: A Filter-and-Refine Approach. In *Advances in Knowledge Discovery and Data Mining*, pages 532–544. Springer Berlin Heidelberg.