# Preface

~~The package also provides functions to post-classification => for post-classification on topics you are interested with. => on topics you are interested in.~~

# Chapter 1

I would add before subsection 1.1 or 1.2 the structure of the chapter (or at least the purpose of the following subsections). During my first reading, I was not sure about the levels of explanation I should expect in this first chapter.

While memory and processors are mentioned for parallel classification, this is not presented for the training of algorithms (trees in RF can be learned in parallel, network parameters can benefit from GPU computing, *etc.*). EDIT: see also my comment on Chapter 6.

Typos

# 1. Introduction

at different times[1]: a space is missing but ~~they~~ may also

BFAST for detecting breaks [9], TIMESAT for extracting phenological attributes [10], CCDC for continuous change detection [11] => for break detection […] for phenological attribute extraction

comes a cost => comes at a cost the *sits* API focus => focuses Parellel => parallel

is *sits* is well suited => in *sits*

to land-related data => for land

# 1.2 Handling Data Cubes in sits

by a start and end date => by a start date and an end date / by start and end dates and cover single date => and cover a single date

1.3.1. **Data structure**

“Soybean-Corn.” => “Soybean-Corn”.

# Sample quality control using clustering

One of the key challenges of machine learning is improving => is to improve

# Classification using machine learning

Caption of Figure 1.5. is in geographic coordinate system (WGS84 {}) => in a geographic […]. I would also remove “{}”.

# Smoothing and Labelling of raster data after classification

Alternatives are gaussian => Gaussian

# Validation techniques

The detailed metrics: the end of the sentence is missing. are available in [Chapter 8](https://e-sensing.github.io/sitsbook/validation-and-accuracy-measurements-in-sits.html): a full stop is missing.

* 1. Final remarks

provides additional function => functions

post-processing and validation: a comma is missing data cubes composed by tiles => of tiles

# Chapter 2

From my understanding, extracting a datacube from multi-tiles is still an ongoing issue.

I feel (the end of) Section 2.1 should be rephrased a bit to answer the following questions:

* + - What is meant by Option 3? (That was not clear to me.)
    - What are the additional files required for Option 2? Do you mean cloud information? If yes, I think cloud information should be seen as an additional band of the datacube, or maybe an extra tibble column in the

data tibble (if this makes sense), but not as an additional file. Having cloud information encoded into the data cube would probably make it easy to apply interpolation techniques (*e.g.*, linear temporal interpolation).

* + - Although it is argued that Option 2 is not feasible (as it would require additional information), the

`sits\_regularize` function is presented as a solution to Option 2 that does not require extra information.

Section 2.2.1 states that *sits* does not crop images when a bounding box is provided to the `sits\_cube` function. Is there any reason not to crop the data (I would have thought that cropping would have improved the memory usage)? Finally, how *sits* manages the overlapping areas (10 km) between two contiguous Sentinel tiles?

Typos

# Image data cubes as the basis for big Earth observation data analysis

defines a compact space[18]: a space is missing its spatial temporal => spatio-temporal

does not correspond to the reality => to reality

The image is: I would have used plural (“Images are […]”)

such as AWS, Microsoft and Digital Earth Africa: a comma is missing An ARD image collection is set of => is a set of

or combined set of sensors => or a combined

where each tile is associated to => is associated with

from the UTM (Universal Transverse Mercator) => from Universal Transverse Mercator (UTM)

Use of tiling system => The use OR Using a tiling system has a valid value, since => value since

Some of images have => Some images These clouds => Values of cloudy pixels tile 20 LLP: there is an extra space

The Brazil Data Cube (BDC) is an example[20].: a space is missing to compose a proper data => proper data

# Accessing Data Cubes and Image Collections in SITS

by STAC (SpatioTemporal Asset Catalogue), by means of the [rstac](http://github.com/brazil-data-cube/rstac) package. [STAC] (<https://stacspec.org/>) is a specification of geospatial information which has => by SpatioTemporal Asset Catalogue (STAC) […] STAC (<https://stacspec.org/>) […] information, which

# Accessing data cubes in Amazon Web Services

Sentinel-2/2A level 2A files : This is not clear to me if “/2A” meant Level 2A (by opposition to level 1C) or the sensor Sentinel-2A (by opposition to its twin Sentinel-2B). I would not use “Sentinel-2A” or S2A (as it is done later in the paragraph) for level 2A, but rather “Sentinel-2 L2A”. Moreover, the tiling system is the same for all Sentinel-2 data (regardless of the sensors 2A or 2B and the correction level).

“B02,” “B03,” “B04,” “B05,” “B06,” “BO7,” B08“,”B8A“,”B11“, and”B12".: some spaces are missing and quote marks are misplaced

All 12 bands are available at 60m resolution: maybe indicate what is the Sentinel-2 spectral band that it is not processed by AWS (B01, B09 or B10?)

sits does does not crop them => *sits* does not crop is an 100x100 km2 => is a

an *sfc* or *sf* object from the *sf* package: it is not explained what is an sf(c) object.

# Accessing the Brazil Data Cube

for all Brazil => for all of Brazil

The large grid is composed by tiles => of tiles The small grid present tiles => presents

The user needs to specific => to specify

# Defining a data cube using files

the parsing info => information

# Regularizing data cubes

large areas, In this case, = > areas. This

which calls the “gdalcubes” package[17]. => package [17].

# Chapter 3

In section 3.5, the sits\_envelope function is not described. Is that “normal”?

Typos

# 3.1 Data structures for satellite time series

at 250-meter spatial resolution => at a 250-meter

“Forest,” “Cerrado,” “Pasture,” “Soybean-fallow,” “Fallow-Cotton,” “Soybean-Cotton,” “Soybean-Corn,” “Soybean- Millet,” and “Soybean-Sunflower”: comma location

This data is also organized as a tibble, with a column with the dates and the other columns with the values for each spectral band.: An example here would be great.

# Time series visualisation

Caption of Figures 3.1 and 3.2. from data set => from a data set (maybe remove it)

# Obtaining time series data from data cubes

Function sits\_get\_data() takes => The sits\_get\_data() function takes In the case of CSV text file => of files

all samples share the same start and end data. => dates Thus, all time series => all of the time series

using the sits\_get\_data. => using sits\_get\_data.

from the attribute file associated to the shapefile => associated with the shapefile

(in code comment) # define the name of attribute of the shapefile => define the attribute name of […]

# Filtering techniques for time series

The sits have support => The *sits* package support Savitzky-Golay

The first two filters are […] while the remaining two have been developed: only three filter are mentioned in the previous sentence

The authors compare => compared

Many other techniques can be found in applications of satellite image time series such as curve fitting [26] => I see Savitzky-Golay or asymmetric Gaussian filters as curve fitting approaches (as first a given curve is fitted to the time series)

# Savitzky–Golay filter

is the the filtered time series => is the filtered

will be equivalent to moving average filter => to the moving average filter each Cj smoothing coefficients => coefficient

of Savitsky-Golay filter => of the Savitsky-Golay filter the Savitsky-Golay filter remove => removes

# Whittaker filter

the effect of Whitakker filter => the Whittaker (missing the and incorrect spelling of Whittaker)

# Chapter 4

The computational complexity is not so much detailed in this section (only one related sentence “In the example below, we take only 10% of the samples for faster processing. Users are encouraged to run the example with the full set of samples.”). I think AHC has at least a complexity in O(n²) with n the number of time series, whereas SOM complexity is linear with respect to n. It might be worse to add one or two sentences on the overall complexity of the proposed approaches (*e.g.*, how long it would take to apply these techniques if I have a dataset composed 1 million labelled instances?).

Opinion: I am not convinced by the experimental setting of the comparison conducted in Section 4.3.3. The OA improvement should be assessed on an independent test set. In the current setting, this is likely that the observed improvement (from 91 % to 95 %) is due to the removal of some “hard” instances ( but maybe not noisy ones).

Typos

sits provides support for two clustering methods to test sample quality, which is => which are

# Hierachical clustering for Sample Quality Control

Title: Hierarchical

# Creating a dendogram

Title: dendrogram

for the purposes of samples data exploration => for sample data exploration criteria are: => criteria are (remove the colon)

# Using a dendrogram to evaluate sample quality

In this example, the height optimizes the ARI and generates 6 clusters. => at this stage, an example has not been provided, maybe “In the following example”

Unlike other validity indexes […] indices => indexes […] indexes OR indices […] indices The contingency table […] help => helps

# Introduction to Self-organizing Maps

a time series of 25 instances => do you mean a time series dataset composed of 25 time series OR a time series of length 25? (not clear to me the use of instances)

a SOM grid is composed of units called .: a word is missing to the same neuron or to nearby ones. => or nearby ones

# Using SOM for removing class noise

Bayesian inference we estimate => The Bayesian inference estimates the neighboring neurons [Santos2021] : citation format (not number)? and (14) Wetlands. => and Wetlands

is a list with 4 tibbles => is a list of 4 tibbles

According to the text, sits\_som\_map returns three tibbles (original time series tibble, neuron, and SOM grid). the SOM grid To => the SOM grid. To

we have a situation the sample is part of => we have a situation where the sample

# Conclusion

this paper proposed => we describe on this book (reference to the paper should be added to Section 4.1)

# Chapter 5

There are several flaws on the RF description:

* + - A random fraction of the features is selected at each node, not at each tree. Each of these features is then tested, the one maximising the decrease in a purity measure is used to build the trees (from the text it looks like several features might be used, this is not the case in the Random Forest algorithm proposed by Breiman in 2001).
    - Variable selection might be performed once all the trees in the forest as been built (not during the training process); The permutation measure is used to evaluate the importance of each feature, not to determine the best split.
    - In the original RF algorithm, trees are built until their maximal depth (*i.e.* when leaves are pure). For computational reason, it is common to use some other stopping criterion (a maximum depth, a minimum node size, a minimum impurity index value, *etc*.)
    - The classification performance depends on the number of trees in the forest as well as the number of features randomly selected at each node.

I would try to keep the terminology consistent through the book. Although previous chapters were using “features”, this chapter often uses “attributes”. I feel features is a correct machine learning term.

Typos

for classification of land use and land cover => for land use and land cover classification deep learning and neural networks. => learning, and neural networks.

# Machine learning classification

Most often, researchers use a approach.: This sounds vague (what is the “approach”?) (and “an approach”)

A common approach is deriving a small set of phenological parameters from vegetation indices, like the beginning, peak, and length of growing season [51], [52], [53], [54].: This is not true for [54] reference (I am the first author here) as we enrich (not derive) initial features with some phenological parameters (and we show that most information is already contained in spectral band features).

These phenological parameters are then fed in specialized classifiers such as TIMESAT [10]: TIMESAT is not a classifier, but a software that allows phenological feature extraction.

It aligns all time series => It align time series

1D convolution neural networks and mixed approaches such as TempCNN [57]. => TempCNN is not a mixed approach, and it is based on 1D convolution. Suggestion: 1D convolution neural networks such as TempCNN [57].

For clarity, I would change sits\_deeplearning to sits\_mlp

“Pasture,”“Soy\_Corn,”: a space is missing

“MIR”) .: there is an extra space

in the R package “sitsdata,” => “sitsdata”,

from the website associated to => associated with

# Visualizing Samples

to the time series associated to the each label => associated with each label provide some insights over => insights into

# Common interface to machine learning and deeplearning models

Title: deep learning (a space is missing)

what labels are have been assigned => what labels have been assigned in [Chapter 6](https://e-sensing.github.io/sitsbook/classification-of-images-in-data-cubes-using-satellite-image-time-series.html) we disscuss => we discuss OR we will discuss

# Random forests

via *bootstrap* procedure => via *bootstrap*

a majority voting schema. => scheme.

improving prediction of => the prediction OR predictions

Random forest training process is => The random forest training process is nondeterministic. => non-deterministic.

the observations set => the observation set the spliting criteria => splitting

given minimum nodes size => node

dependent on the number of decision trees to grow and of the “importance” parameter => dependent […] on the

“importance” parameter

I would have used sits\_rf (rather than rfor) as RF is the common acronym for Random Forests the randomForest package[59]: a space is missing

its main parameters is num\_trees => parameter Random forest classifier are => Random forests are random forest tend => forests

# Support Vector Machines

are linear (p−1)-dimensional boundaries: p variable has not been introduced only on those samples that violates => violate

All other points far away from the hyperplane does not => do not on the hyperplane coefficients which let SVM => coefficients, which

generally translates its hyperplane to a nonlinear boundaries => boundary The use of kernels are => is

an hence => and hence

as a wrapper of e1071 R => as a wrapper of the e1071 R [61], sits adopts => [61]. The sits package adopts

is not a robust to outliers as => is not as robust. This is a bit counter-intuitive as the introductory paragraph explained that SVM was robust to outliers. The issue here is that the outlier notion is used in two different contexts: outliers as an extreme value within the time series (the context here), and an outlier as an observation that has a different behaviour to other observations (in the subsection introduction).

Figure: caption is cut. And what is “ands”?

The result is mostly consistent of => English? Maybe use “is coherent with”? does not contain a samples => a sample

places where forest is removed will => places, where the forest is removed, will for year 2009 => for the year 2009

cost of contraints violation => regularization hyperparameter OR generalizing hyperparameter (and constraints)

Such sensitity => sensitivity

from 100 to 1 […] Such sensitity to the input parameters points to a limitation: It should be mentioned here that it is current to use a k-fold cross validation procedure to tune the hyperparameter values of SVM. Furthermore, why values “from 100 to 1” for the cost hyperparameter?

strinkgly => strikingly

# Extreme Gradient Boosting

by fitting better model => a better model to build a better prediction => predictor

the same random logic for tree selections is applied at every step.: I do not understand this sentence. I think

“selections” is not correct in this sentence. to improve on previous result => results

at the risk of overfitting the model. => at the risk of overfitting.

minimized => minimised (British English spelling seems to be generally used) by agradient: a space is missing

The minimim loss => minimum how deep tress are => trees is

it should not be largem => be too large

higher depth trees lead => high tree depth leads to The nrounds parameters => parameter

However, for each specific study, users need to perform validation.: Maybe explain why here (needs to tune hyperparameters).

# Deep learning using multi-layer perceptrons

Using the keras package [63] => Keras (several times)

SITS supports the following =deep learning techniques, as described in this section and the next ones => remove the

“=”, and what about one section on deep learning and then subsections about the different approaches? a multilayer perceptrons => perceptron

for a classifier y=f(x) maps: remove “for”

by the user, since => by the user since

The most important decision is the number of layers in the model.: and also the number of units/neurons per layer and test different options of number of neurons per layer => and test different numbers of neurons

Three other important parameters for an MLP are: (a) the activation function; (b) the optimization method; (c) the dropout rate. => I would not consider the dropout rate has a main parameter of MLP, as many MLP architectures does not use it. I would present it as way to reduce the risk of overfitting (as the weight decay for example).

The activation function the activation function: remove one expression are gradient descent algorithm => algorithms

this procedures prevents => procedure

using an example of the deep learning method. => using an MLP model. for input at a given steps; => step;

randomly set side => randomly set aside

To simplify the vignette generation: The “vignette” has not been explained (yet). this series to class => to a class

shows an increased sensitivity to the data variation than the previous models=> an increase in the sensitivity to data variations compared to previous models

# Combined 1D CNN and multi-layer perceptron networks

similar to the fullCNN model discussed above,: not anymore the case (see figure). => (see figure 5.1).

controls the number of 1D-CNN layers => and the corresponding number of neurons in each layer We recommend to set => We recommend setting

While the result of the TempCNN model using default parameters is similar to that of the MLP model => It does not seem exactly the case with the display results. Maybe explained that the overall results are similar (or maybe this statement is not valid anymore after the code correction?!)

We recommend to set => We recommend setting it has has the potential: remove one “has”

than the the MLP model: remove one “the”

this models => these models

# Residual 1D CNN Networks (ResNet)

is divided in => divided into The idea is avoid => is to avoid

based gradient optimization methods[70] : a space is missing + based on that has shown => that has been shown

the size the of kernels of => the size of kernel Curve caption is cut (classification example) in year 2004 => in the year 2004 OR in 2004

such labeling is useful to point out to missing classes => to point out missing classes in the area started in 2008. => 2010(?)

# Chapter 6

I would clearly make a distinction between (i) training, and (ii) prediction. I feel “classification” is used in this chapter interchangeably for both cases. Hence, it can be explained that the training of sits\_rfor and sits\_svm can ba done by using parallel process, but this is not the case for other algorithms.

Although CPU and GPU usage are not mentioned here for deep learning algorithms: can the user chose to run the training and/or the prediction on CPU or GPU?

Also, I would indicate the machine used for experiments in section 6.4.

Typos

are associated to => are associated with

# Data cube classification

To classify a data cube, use the function sits\_classify(). => one can use OR the sits\_classify() function can be used.

The classification algorithms allows : plural or singular ow many process will => processes

The code bellow => below

a small raster brick image that accompany => accompanies (maybe “that is part of the package”. Also, the “brick image” notion is not clear to me.)

Once a data cube which has associated files => cube, which. What do you mean by “associated files”? such disk read speed => such as disk read speed

the classification task it is split: remove “it”

The underlying algorithms available in these models already have parallel processing facilities. => Are “these models” deep learning ones and xgboost? I don’t think xgboost has that much parallel processing facilities (I might be wrong though).

# Processing time estimates

from experiments made the authors => made by the authors

with a random forest model with 2000 trees: Comment: 2000 trees seem a lot.

# Chapter 7

Typos

# Introduction

method to remove => in order to (as “to” is already used earlier in the sentence) bayesian smoothing => Bayesian

gaussian smoothing => Gaussian

smoothing and bilinear => smoothing, and bilinear. (Is bilinear smoothing described in this chapter?)

, as shown in the examples below: there is no example shown.

# Bayesian smoothing

an be thought of as way of => as a way (and this is maybe too wordy) probability is => probabilities are (OR the probability is)

The distribution π(x|θ), called the ,: a word is missing (“likelihood”?)

Bayes’s rule states that the probability is proportional to the product of the and the probability.: many words are missing here

which is the random variable that describe => describes

In the case of smoothing of image classification => of smoothing an image classification In this section, the x variable is not defined (for the given context)

# Derivation of bayesian parameters for spatiotemporal smoothing

We label a pixel pi as being of class k if => of class k at time t if

where pi,t,k is the probability of class k at time t. => where pi,t,k is the probability that pixel I belongs to the class k at time t.

xi,t=(xi,t,k0,xi,t,k1,…,xi,t,kK): indices from 0 to k-1 OR from 1 to K This formulation allows uses to use => allows us to use

which measures the *logit* (log of the odds) associated to classifying => English?

by applying the density functions associated to the above equations => associated with then our confidence on the influence => in the influence

* 1. **Use of Bayesian smoothing in SITS** Section number: should not be section 7.2.2? is set of probability images => is a set

Caption of Figure 7.2: Probability values smoothed by bayesian method => by a Bayesian method

some of local variability associated to misclassified pixels which are => some of the local variability associated with misclassified pixels, which

is best appreciated comparing => is best appreciated by comparing

# Bilateral smoothing

Title: I think this bilateral smoothing is introduced in 7.1 as the Gaussian smoothing a pixel with high probability => with a high

Because of its design, bilateral filter => the bilateral filter the values of pixels in a if neighborhood: English?

∥I(xk)−I(xi)∥ is the absolute difference => if it is an absolute difference, || should be replaced by | (otherwise ||.||is

the norm)

For every pixel, the method takes a considers => the method considers Big difference between pixel values reduce => differences

Big distance between pixels also reduce => distances The achieve a satisfactory result, => To achieve between 4 and 16[75] : a space is missing

Caption of Figure 7.5: by bilateral filter => by a bilateral filter

# Chapter 8

Typos

# Validation techniques

the prediction error associated to some model. => associated with some models

The so-called *k-fold* validation, we split the => In the […] OR The so-called […] splits the

into k partitions with approximately => of approximately and proceed by fitting => proceeds

there is a bias-variance trade-off in choice => in the choice This can be computational expensive => computationally the same number of fitting process => processes

gives support the k-fold validation => gives support to the k-fold validation an example on how to => of how to

how to proceed a k-fold cross-validation => proceed k-fold It perform a five-fold => performs

using SVM classification model as => using SVM OR using a SVM classification model

# Comparing different machine learning methods using k-fold validation

we show only the result of TempCNN, that has => , which has OR without the comma Caption of Figure 8.1: cross validation => cross-validation

8.3.2 **Classified images**

For this reason, [76] argue => argues

when measuring accuracy of classified images => when measuring the accuracy