Major comments:

the authors cover many important topics regarding reference data for large area mapping of tree species and supply a valuable dataset for the research community.
While the dataset will lack behind expectation with regard to scientific freedom and flexibility, its provision will provide researchers, educators and students with additional data for the investigation of important research questions related to climate change, forest preservation, forest management and biodiversity studies.
While some aspects of the data set, such as preprocessing, could be up for discussion, its publication will serve as a baseline for future publications of state and federal data sets and hopefully motivate more government authorities to provide their inventory data to the public.
The writing style is excellent throughout the paper with only a single recommendation from my side.
The researchers worked thoroughly on assuring high data quality and investigated the data set at hand for important characteristics, such as distinguishability of species from spectral signatures and geolocation accuracy, something that is to be expected from future related research.
In my opinion, a few aspects of possible data usage were missed in the study design and discussion but overall, the state of the art in the field of tree species mapping with multi spectral satellite imagery is presented correctly.

Minor comments:

106-107: "if their crown is overlapped by less than 50% by the surrounding trees" would be clearer language IMO

2.3:
Additional TSA-processing withing the FORCE framework is not possible and the opportunity to create a dataset of even higher quality is missed. This would undoubtedly lower the amount of available tree observations but might help classification approaches that are sensitive to noise from cloud shadows and fog.

134-135: calculating the area-weighted average of a pixels might be a big source of noise if, let's say, the other 75% of that pixel depict a substantially different type of land cover than the target tree species. Think of the spectral signature of a deciduous tree that is added to a coniferous evergreen and the undergrowth signal in winter observations.
In my opinion, some sort of outlier detection should be put in place to detect possible addition of noise.

276: This might be due to *Pinus*' often very top-heavy crown in plantations that allows undergrowth to be more visible. In combination with *Betulas* characteristic bark, it is no wonder that the signal gets mixed up. There might be similar issues with stands including *Larix* or *Fraxinus*.

310: One additional idea for use of your dataset could be the investigation of mixed pixels.
For large area mapping it would be great to know if any given mixture of species within a single pixel can be learned by a classifier.

On the discussion of pixel level vs tree level reference data:

While I see possible issues with area-weighted pixel extraction, I do not agree with the criticism stated by Reviewer 2:

FORCE uses the ImproPhe algorithm (Frantz 2016) that alters the pixel values of the 20m bands. To my understanding, duplicate values within the vicinity of a datapoint and thus spatial autocorrelation will be quite unlikely given the large size of the dataset. I can also support the author's claim, that duplicates (as well as random noise) within a certain threshold as stated in regard to the LAION 400M dataset are no issues for modern machine learning algorithms, especially neural networks, from my personal experience.

The addition of random noise is a valid point of criticism. However, as long as European NFI rely on fixed position sample plots, this approach seems to be the only viable method to provide data to the research community.