Review of the Article

"Potential of Convolutional Neural Networks for Forest mapping using Sentinel-1 Interferometric Short-Time-Series"

General assessment: This article shows a methodology for mapping forest in the Amazonian basin using short time series of Sentinel-1 in a Convolutional Neural Network (CNN) framework. In particular, it investigates the estimation of temporal decorrelation parameters using the CNN framework instead of existing computing intensive workflows. The approach shows better results than previous state-of-the-art frameworks, while diminishing the computational expense. The article is well written and necessary validation is performed and discussed. In general, I recommend this article for publication after a few clarifications and minor changes.

Specific comments:

- You could already mention in the abstract how short your STS are, as you mention at the end that you get good results using a temporal baseline of 12 days – it could allow the reader to estimate how much scenes, and which temporal baselines, you considered.
- l.40: it is not very clear here if the PRODES, DETER and TerraClass systems are monitoring systems or ongoing projects – maybe you could specify it a bit.
- At the end of the introduction, you should mention and highlight more specifically the novelty of your work compared to previous works, besides its applicability to 12 days interferometric coherence- has the U-Net architecture been used for similar cases before? If not, you should definitely highlight it as a novelty of your work. Alternatively, you should specify that your goal was mainly to enhance an existing approach using S-o-A CNN architectures.
- In Table 2, the corner coordinates of the different stacks are not necessary: the considered scenes and their respective extents are well visible in Figure 1 already (You should add a coordinate grid on Figure 1 though). Instead of the coordinates in Table 2, you could give other important parameters, such as the max/min perpendicular Baseline (- and maybe the mean multi-temporal coherence over the stacks, possibly for the different temporal baselines, or another relevant parameter for the approach)
- The workflow of Figure 2 is very simple and to my point of view should be modified according to the corresponding text. Although the text of section 2.2. is well explained, it does not really corresponds to what is shown on the graphic of Fig. 2. Consider adding in Fig. 2 specific parameters that you calculated for each branch (e.g. gamma nought, multi-temporal gamma nought mean, coherence for different temporal baselines). I'll suggest here to modify Fig.2 to keep only what you've done and not show what the method of [2] was – here the reference should be enough. Add some details about which variables you calculated and if you want to keep the method of [2] as comparison, you could use colors (e.g. light grey) to show that they are no longer used in your current workflow. In general, I would also suggest moving section 2.2. into the 3rd chapter as it rather belongs to the methodology - or you could rename it "data preparation".
- l.140-142: Here, you mention that you generated a database in order to produce monthly reports and regularly update the rainforest maps. This is also a great novelty that you could already mention in the introduction. However, the presented data shows only an example for one particular month. It would be good to discuss a bit more how this could be implemented in a monthly updated systems and what are the remaining challenges (e.g. seasonality, phenology etc…)
- l.171: why did you use a 5x19 moving average? Could you justify it a bit?

- What influence does speckle have on your results? Does it play a negligible role as you are considering patches instead of only pixels as compared to the RF classifier?
- The computational load is described as being a main issue of the methods presented in [1,2] – would it be possible to add an estimation of it and compare it to your proposed approach? (Possibly in Table 4 as additional column – if possible using comparable computing resources)
- In Fig. 7, it seems that small roads are not well classified – is it rather due to the patch learning or to the underrepresentation of impervious areas? Please comment on this shortly in the discussion – as I understand it cannot be due to the considered temporal baseline as for case X, all baselines were used.
- Fig.9 is interesting but I am not quite sure how to interpret it – in the case of Fig.9 (a) it would seem that the water class is therefore overpredicted using model X, right? I think it cannot be directly compared to Fig. 9b, as Fig. 9a would get a higher weight for the water class as for Fig 9b through the normalization and may therefore be overrepresented. Maybe you could add a comment on that that would help further interpretation?

Minor suggestions:

- l.12: there should be no reference in the abstract

- l.77: "in [20] TanDEM-X InSAR data was used to generate a forest map over the state of Pennsylvania, USA". → "TanDEM-X InSAR data was used to generate a forest map over the state of Pennsylvania, USA [20]"

- l.146: "(superimposed over an aerial scene from Google Earth)": you could only mention the superposition on Google imagery in the figure caption – it is not imperative in the text

- l.436: "Moreover, we can see in this case that 19%"→ "Moreover, we can see in the first case that 19%".

- l.469: "it becomes clear the role of …"→ "the role of… becomes clear"