Response to comments by Referee #1

We thank the referee for taking the time to review our manuscript and for providing these useful comments. In this document, we respond to the remaining questions and comments and how we addressed them (referee comments in italic). When we refer to line numbers, we mean those in the revised manuscript.

This manuscript presents a RGB photo dataset collected from a group of near surface Phenocam installed over high Arctic valley. The dataset contains three-years time series of photos with 4-6 hours temporal resolution, and is used as proxy of NDVI indice to monitor vegetation canopies. It concluded that ordinary RGB cameras are a promising tool to identify temporal and spatial patterns in vegetation productivity and composition at a landscape scale, and that GCC compares well to NDVI at the plot level and shows a similar temporal pattern. The dataset is well described , and the methods for data collection and processing are clearly presented. However, some issues need to be handled before the manuscript can be accepted for publication.

We are glad to hear the overall positive impression the referee had of our paper. We respond to each issue down below:

Details:
1) The case study area, a valley of Adventdalen on the Svalbard archipelagois relatively small and may not well represent large areas in the Arctic. Suggest to add some discussions on how this case study can be somehow scaled up to larger areas or applied in similar landscapes.

This is a common discussion when it comes to studies on Svalbard: how representative is it? We have therefore added the following text to the manuscript (line 107), and included the relevant vegetation classes in Table 2:

> The vegetation types at our measurement locations are relatively common to Svalbard but also the rest of the Arctic. In Table 2 we show that our plots cover 6 (out of 11) vegetation classes defined for Svalbard (Johansen et al. 2012), and correspond to three classes of the Circumpolar Arctic Vegetation Map (Walker et al. 2005; Raynolds et al. 2019). These are: Sedge/grass, moss wetland vegetation (W1), Graminoid, prostrate dwarf-shrub, forb tundra vegetation (G2), and Prostrate/Hemiprostrate dwarf-shrub tundra (P2). Furthermore, our plots show strong similarities to two more vegetation classes: Rush/grass, forb, cryptogam tundra (G1) and Prostrate dwarf shrub, herb tundra (P1). Combined, these vegetation classes cover nearly a quarter of the unglaciated parts of the Arctic, mostly in Greenland and the Canadian Arctic Archipelago, but also the northernmost parts of Alaska and Russia. This underscores the relevance of this data to studies of arctic change. In addition, the techniques presented here are applicable to any short stature vegetation type – including grasslands, heaths, croplands and wetlands across the world.

2) GCC is an index composed of RGB bands of digital photo, while NDVI is calculated with near-infrared and red bands. They may have a high correlation in some cases, but we have to make sure such correlation is consistent across space and time. Please add paragraphs to discuss the issue.

We agree that this may be an issue and we already discuss this in section 3. However, since this setup is used to identify phenological changes, we are mostly interested in the direction of change and less in the absolute value of the remotely sensed signal. The timing of phenological stages compares quite well, and this has previously been concluded by others as well (see e.g., Richardson et al., 2018; Sonnentag et al., 2012). But to further highlight this issue, we have expanded the last paragraph of section 3 which now reads as follows:

> Another issue arises from the differences in the spatial patterns between NDVI and GCC, which may make it challenging to compare vegetation maps that are based on either index. Since NDVI

> relies not just on the visible part of the electromagnetic spectrum but also the near infrared, it would be expected that differences in the amplitude of these signals arise when responses in the visible and near-infrared bands diverge. However, the main purpose of this dataset is to assess the timing and pattern of phenological stages which are derived from the direction of change in either NDVI or GCC, rather than the absolute magnitude of these indices. Figure 5 shows that the pattern of green-up, peak growing season and senescence compare quite well, as found previously by others (e.g., Richardson et al. 2018; Sonnentag et al. 2012), and we expect that the application of NDVI and GCC to assess phenological timing will be relatively similar across the landscape. Therefore, we consider the two vegetation indices as complementary. While NDVI responds to increased overall growth of vegetation (reflected near-infrared light and absorption of red light), GCC responds to the changing level of green pigments in the vegetation.

Also, we address the possible issue of misalignment of the NDVI sensor and RGB cameras on line 330:

> While considering spatial and temporal differences, it appears that the relationship between GCC and NDVI is rather consistent from year to year when the same plot is considered. The possible exceptions are plot 2 and 4, which contained a large fraction of bare ground. When the RGB camera and NDVI sensor are not pointing at the exact same area in such heterogeneous landscapes, the amount of bare ground and vegetation in their field of view will diverge, causing a relative difference in magnitude between the two indices.

3) Page 3, line 80-81: Please explain why the dataset ends in 2018, any plan to continue and update?

The data is still being collected but due to equipment failure no data was collected in 2019. The data from 2020 and onwards will be processed at a later time and added to the dataset. We added the following sentence to the last paragraph of the introduction (line 89):

> This dataset will be updated with data from following years (2020 onwards) according to the protocol laid out in this paper.

4) RGB photos were captured with several different models of digital camera, from GardenWatchCam, WingScapes TimeLapseCam, to Breinosa. Suggest to discuss how these cameras are cross-calibrated and how the photos are normalized to ensure consistency.

We ran a GardenWatchCam and WingScapes in parallel on rig #1 for two weeks in 2017. This showed that, on average, the two had similar values for GCC but that the GardenWatchCam showed more variation (i.e., noise). This general behavior is mentioned in section 2.3.2. However, just like the potential differences with NDVI, this setup was designed to track the phenology of the vegetation, and we do not expect the timing of those stages to be different with different cameras. This is also an argument made by Richardson et al. (2019), who primarily pointed to automatic white balancing as a potential issue that could affect measurements. Therefore, we have added a new figure showing the comparison (Figure 4) and added the following sentence to section 2.2.1 (line 132):

> Both camera types were used at their highest image quality setting, with default settings that do not include automatic white balancing since this has been pointed out as essential to achieve a consistent sensor response (Richardson et al., 2019).

And to the end of section 2.2.2 (line 169):

> No automatic white balancing was used on any of these cameras.

And the following to section 2.3.2 (line 246):

> We determined whether these greenness indices differed between camera types (GardenWatch and WingScapes) by operating these cameras in parallel on rack 1 for a few weeks in 2017. This showed that GRVI differed quite strongly, while GCC was highly consistent between camera types (Figure 4). Since this index also showed lower variance and correlated best with NDVI, when considering all plots, we use GCC throughout this manuscript.

5) Page 5, line 145-146: I don't quite understand the sentence "... the blue channel had been replaced with a sensor sensitive to the near infrared, ...". Is it a near-infrared band or just somehow sensitive to near-infrared? Please clarify.

It's a near-infrared band (sensitive up to 920 nm). This has been changed in the text.

6) Page 5, line 155-156: I would suggest that we keep the photos that contains snow on nthe ground, as snow is such an important feature in the Arctic and such photos may be useful for research.

Because of ease of use of this dataset, we don't retain photos with snow for the small rigs, also because it interfered with the stabilization algorithm. However, for the landscape photos we did retain those with snow cover visible, precisely because snow is such an important feature, and this shows how snowmelt differs across the landscape. This is now mentioned in the text (line 177) as follows:

> For these photos, images with snow on the ground were retained to show how snowmelt differs across the landscape.

7) Page 9, line 263-264: The major difference between NDVI and GCC is contributed by the contrast between near infrared and blue bands. Here you need little in-depth discussion on how they differ and therefore influence the correlation between NDVI and GCC in general.

See our answer to comment #2.

8) Page 10, line 311-313: With a data format of JPEG images, I wonder how spatial coordinates are provided in the datasets, in metadata? Please clarify.

Each timeseries of photos is accompanied with a textfile that contains all metadata. This has now been specified in the text (line 345).