SCIENCE RESULTS FROM THE PLANETFOUR CITIZEN SCIENCE PROJECT

K.-M. Aye[1], M. E. Schwamb[2], G. Portyankina[1], C. J. Hansen[3], [1]LASP, University of Colorado at Boulder, Michael.aye@lasp.colorado.edu, [2]Institute of Astronomy & Astrophysics, Academia Sinica (ASIAA), Taipei, [31]Planetary Science Institute, 1700 E. Fort Lowell, Suite 106, Tucson, AZ 85719

Introduction: PlanetFour is a Citizen Science project about analyzing surface images from the south pole of Mars (www.planetfour.org). Main objectives are studying the surface atmosphere interactions at the pole, especially during the local spring. CO_2 gas jets that are created by basal sublimation of the seasonal CO_2 ice layer deposit fine dust into the atmosphere and coarser regolith on top of the ice sheet in form of fan-shaped albedo features [1]. The fine dust that entered the atmosphere is believed to have an important effect on the atmospheric temperature profile. The seasonal removal of regolith over many years results in topographical features called araneiform. These are dendritic troughs that connect to a common center. Their constant modification represents ongoing change in the surface topography of Mars today [2].

A further objective is to map the orientations of the regolith deposits. These orientations are controlled by the local winds that existed at the time of jet eruption. Repeated surface observations constrain the time of eruption and are therefore able to provide wind data points for atmospheric meso-scale simulations.

Data: The image data used in the PlanetFour project comes from the HiRISE camera of the Mars Reconnaissance Orbiter. Planet Four citizens are asked to identify and outline fans in the presented tiles that are created by tiling up each large HiRISE observation in screen-sized tiles of approx. 800x600 pixel. We cluster the resulting markings into final locations using the DBSCAN [3] clustering algorithm, after which the object coordinates are back-projected into latitude and longitude coordinates with a standard ISIS image calibration pipeline.

Outcomes are catalogs of object locations, estimated sizes and wind directions. With these catalogs we study the activity over time per region, compare these activity time-series between seasons and compare the strength of observed activity between different regions around the pole. The derived wind directions are used to improve atmospheric meso-scale simulations at Mars' south pole, which is part of an ongoing NASA SSW project. We will present results from the first publication of the PlanetFour project.

Methods: Using DBSCAN (as provided by the scikit-learn package [4]) we first cluster on the simple x,y position of the respective marking. Because we encouraged the volunteers to make multiple markings from the same point for fans that seem to reappear but deposit into a different direction, we then need to cluster on angles, to differentiate between these. The controlling parameters for DBSCAN are the maximum distance allowed in the currently used metric, called "eps' and the minimum number "min_samples" of markings to define a cluster. For example, in pixel space an eps value of 10 would mean that only markings within 10 pixel can be linked into a cluster. Additionally, markings that only can be linked once are marked as non-core members that we filter against for higher data fidelity.

The angular clustering needs to performed in rectangular space to compensate for the circularity of angular values. For this we simple calculate the polar coordinates for a given angle. In case of the ellipse markings for blotches, we then only cluster on the sin(angle) due to the inherent symmetry of ellipses, while we use cos(angle) and sin(angle) for fans.

Once we have determined which markings make up a cluster, we create an average object for an intermediate planet4 tile-based catalog. Now we combine all the tile clusters per HiRISE observation, checking for overlap in determined clusters and once again merging them if they fulfil a minimum distance criteria. (PlanetFour.org serves HiRISE tiles with 100 pixel overlap for best marking efficacy). In case of undecided votes of blotches against fans, we write down the voting ratio, and create a default catalog that makes this decision at the 0.5 ratio point. We expect that the end-user would adapt this ratio for filtering the catalog depending on the science use case. This end product catalog will be sorted directly by HiRISE observation ID, due to those determining the location and time of available data. But we will produce an intermediate tile clustering catalog as well for the end user that requires more insight.

Discussion: A lot of efforts was spent in getting the clustering right. Due to the massive amount of available data for PlanetFour, we decided to reduce the number of required markings to a level that enables us to retire more PlanetFour tiles in a reasonable time. However, it turned out that our retirement criterion of 30 data points per tile is already in a realm where it is more difficult to render stable results than standard clustering applications usually apply. (Hundreds of samples instead of dozens). We tried to apply a more modern density-based clustering algorithm called HDBSCAN. It is a further development on DBSCAN done by partially the same team that initiated the DBSCAN algorithm[5]. One of its advantages is that it can deal with multi-dimensional scattering more easily, enabling to include angles and

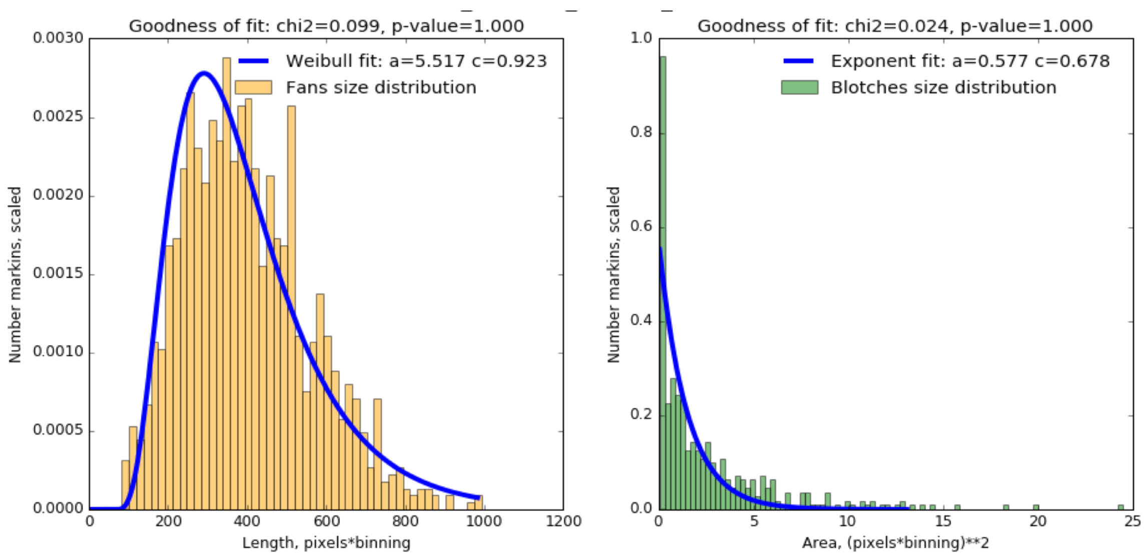

Figure 1 Example of PlanetFour science pipeline output. All fan and blotch markings for HiRISE image ESP_011403_0945 have been scaled for the observation binning mode (to correct for the covered ground area). Wind related phenomena often can be fit well with a Weibull distribution. As expected, fans fit much better into this distribution, as they are believed to be created by the prevailing winds during the jet eruption. Blotches, however, are believed to have their shape due to lack of winds during the eruption. Therefore this analysis is studying different things: On the left we have a study for the prevailing wind strengths with a fan size being a proxy to the wind strengths, while the right hand side is mostly studying the statistics of the eruption strength, because without wind the eruption strength becomes the dominant physical parameter determining the blotch size.

radii into the process. But it turned out that our data densities were too low to render the clustering stable.

We are now back at using DBSCAN, but clustering sequentially, first on locations, then on angle, and potentially on radii and fan spreading angles. While this makes the code more cumbersome, the results are easier to understand and better to control than the all-in-one solution offered by HDBSCAN.

Science analysis: We have established a first science pipeline that takes a set of final clusters per HiRISE image, scales the marking sizes by the HiRISEbinning observation mode (covering different ground areas), and then fitting a Weibull distribution over these. Wind related phenomena often fit well to a Weibull plot, and one can see in Figure 1, how the fan sizes can be fit well, due to them being deposited into fan shapes by the prevailing wind during a CO_2 gas eruption.

We have created these plots for the available seasonal monitoring HiRISE images per location and at different location and are now able to quantify differences in local eruption activity, wind strengths and wind variability and will present our recent results spanning several years and south polar locations at the conference.

References: [1] Kieffer, H.H., 2007. Cold jets in the Martian polar caps. *Journal of Geophysical Research* 112, 08005. doi:10.1029/2006JE002816 [2] Piqueux, S., Christensen, P.R., 2008. North and south subice gas flow and venting of the seasonal caps of Mars: A major geomorphological agent. Journal of Geophysical Research 113, 6005. doi:10.1029/2007JE003009 [3] Ester, M., H. P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, pp. 226-231. 1996 [4] Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research,* 12, pp. 2825-2830, 2011. [5] K. Chaudhuri and S. Dasgupta. *"Rates of convergence for the cluster tree."* In Advances in Neural Information Processing Systems, 2010.

Acknowledgements: This work is partially funded by NASA SSW grant no. NNX15AH36G.