CHANGE DETECTION IN REPEAT IMAGERY USING PRINCIPLE COMPONENT ANALYSIS. Reid A. Parsons[1,2] and H. Miyamoto[3] [1]Department of Museum Collection Utilization Studies, University of Tokyo, Japan (rparson4@fitchburgstate.edu), [2]Earth and Geographic Sciences, Fitchburg State University, U.S.A. [3]Affiliation Department of Systems Innovation, School of Engineering, University of Tokyo, Japan.

Introduction: Change detection is an analytical tool of Remote Sensing in which repeat observations of a particular location can be overlain, subtracted, or ratioed to detect differences between images. High-resolution, repeat imagery on Mars has led to several discoveries of ongoing geologic processes operating at the sub-kilometer scale such as recurring slope lineae (RSL) [1-3], new, small impacts [4], aeolian dune migration [5], sublimation of polar ice [6], and gully formation [7]. Generally, the detection of these changes is done manually by overlaying, and "flipping" between the two images. As planetary remote sensing data sets continue to grow, so does the need for a more efficient means of change detection.

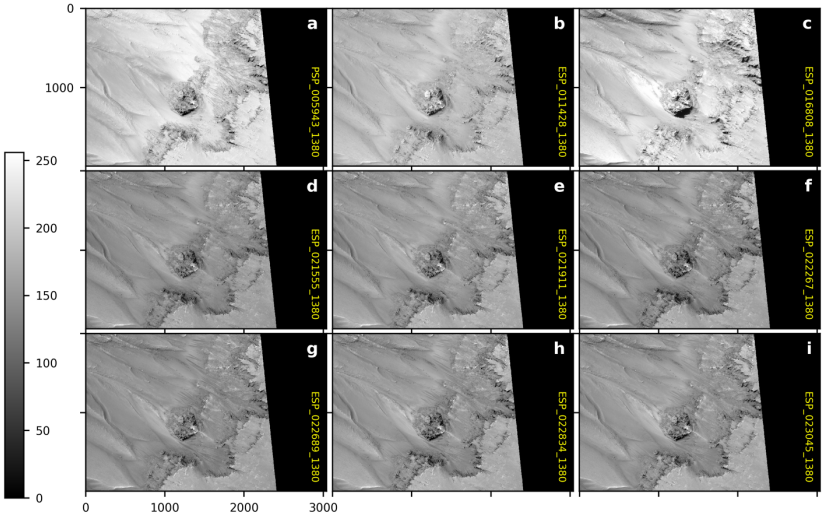

Fig. 1. HiRISE time-series image stack of the southeaster wall of the 16 km-diameter Palikir Crater (41.65S, 202.71E), Mars.

Principle component analysis (PCA) offers a advantageous means for detecting change because it considers the values of all pixels in all the repeat images to extract the most statistically significant variations. These variations are then isolated from one another as a hierarchical set of "principle components" of the image dataset. The user can quickly identify patterns or features of interest which are present in one or more of these components. To illustrate this method, nine HiRISE images [8] collected from Palikir crater on Mars between November 2007 and June 2011 documenting RSL formation (Fig. 1) will be used. The method is herein referred to as "change detection via PCA of stacked time-series" or CDPCAST.

Generating a Time-series Stack: Preprocessing involves image calibration (converting camera data into reflectance; optional if same camera is used in all images), orthorectification (removing geometric distortion associated with viewing angle and topography), and co-registration (aligning images of the same location so that they perfectly overlap). In addition to these initial steps, CDPCAST requires that each image in a time-series be of the same extent and resolution and aligned (via coregistration) so that they can be "stacked" into a single, time-series image file. Fig. 1 shows each of the preprocessed, uncalibrated HiRISE images in the final image stack with source image ID given in yellow. Axis labels indicate numbers of pixels, and the color bar indicates the pixel values. Each image has a resolution of 0.25 m/pixel.

PCA: PCA is a common data reduction method employed in the field of remote sensing, but is most often applied to multi- or hyper-spectral images in order to distinguish between different materials in an image by accentuating differences in reflectivity at different wavelengths of light. The CDPCAST method, in contrast, performs a PCA on a stack of images taken at the same wavelength at many different times. Although prior studies have employed PCA as a means of change detection on Earth [11-13], the remote sensing community has focused more on alternative methods [14].

PCA changes the values of the pixels in each image by plotting the pixel data on a set of axes which capture the maximum variance in the dataset. In our example dataset, each pixel in the image stack can be treated as a 9-dimensional vector since there is a value for that pixel in each of the nine images. However, because many pixels experience little change between images, there is redundancy in the stacked image data that can be reduced by plotting the pixel data on a new set of axes that capture the most important, or PCs of the image stack.

Results: Fig. 2 shows the result of the PC transform (performed using ENVI® software) on the stacked time-series. Because most of the variation in the images is captured in the first PC (PCA1), it is colored according to the color bar given in Fig. 1. Generally, variations covering a large spatial extent or variations causing a big shift in pixel values will be captured in the lower PCA components. Higher components, on the other hand, will depict smaller amplitude variations occurring in fewer images (or over smaller spatial extents). Thus, components 2 and 3 (Fig. 2b and 2c, respectively) are dominated by variations caused by differences in lighting (sun direction and elevation angle) and possibly dust

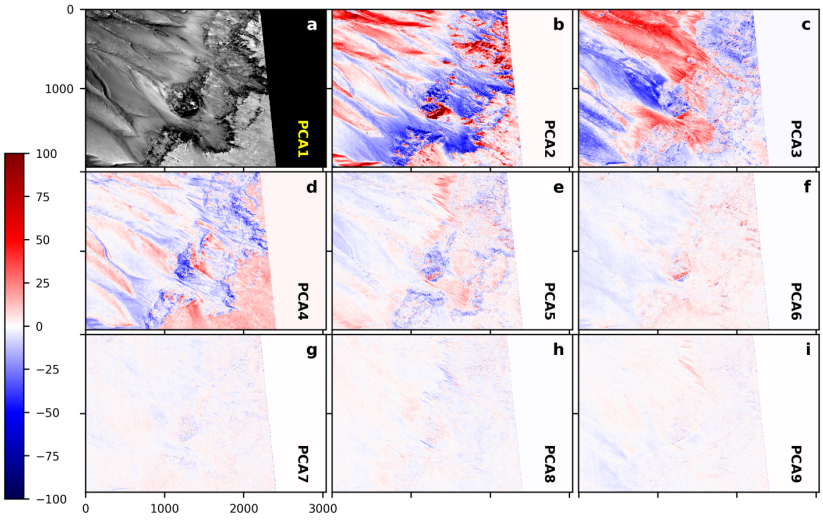

Fig. 2. Nine components of the PCA transformation of the image time-stack. PCA1 (a) uses the gray color scale used in Fig. 1 due to its large variation in pixel values. Components 4, 5, 8, & 9 (d, e, h, & i) contain RSL.

or frost deposition/erosion. RSL features are identifiable in PCA4 (Fig. 2d) as elongated blue "fingers" and become more dominant and appear red in PCA5 (Fig. 2e). Subsequent components capture smaller variations caused either by changes in lighting, RSL shape (in PCA 8 and 9), or camera noise.

Mapping PCA Features to Original Images: Having identified PCs 4, 5, 8, and 9 as containing the feature of interest, the next objective is to identify which of the original images contain RSL. While this task is trivial for a small number of images, ongoing data collection may result in dozens of images in certain locations [15]. Before image attribution can be performed, the user must first construct a "feature vector" (f) whose elements (=0 if feature is not present, =1 if feature is present, =-1 if feature is present, but pixel values inverted relative to the PC in which the feature is first found) correspond to each of the PCs. Because PCA1 is the average image of the dataset, it is the reference from which changes are revealed and will always be set to zero. Once the feature vector has been determined, we can quantify the potential (p) of an original image containing the feature via the following equation: $p = Ef\sqrt{\lambda}$ where E is the matrix whose columns contain the orientation of the PC axes in image data coordinates and $\sqrt{\lambda}$ is a vector giving the variance of the image data along each PC.

In practice, determining the elements of f can be ambiguous. For example, RSL in PCA8 (Fig. 2h) appear blue, which is the same coloration of RSL in PCA4 (the first PC in which RSL are distinct), so the 8th element of f should be 1. Solving for p in this instance (solid line in inset of Fig. 3) results in images (a), (g), (h), and (i) (see Fig. 3) having the highest probability of containing RSL. While this is a successful result, image (f) also contains small, insipient RSL, but is given a low probability. Closer inspection of PCA8 reveals that small, insipient RSL are colored red instead of blue – suggesting $f[8]$=-1. Updating the calculation of p (dotted line in Fig. 3 inset) gives a more favorable probability for image (f).

Due to uncertainty in determining the presence of the feature in a given PC, determining which images in the time-series stack that are responsible for the variations revealed in the PCA is the potentially unreliable part of the CDPCAST method and requires some caution. However, the method is reliable in predicting which images contain the most prominent RSL.

References: [1] McEwen A. S., et al. (2011) Science 333 740–743 [2] Sullivan R., et al. (2001) JGR 106 23607–23633 [3] Ojha L., et al. (2014) Icarus 231 365 – 376 [4] Daubar I., et al. (2013) Icarus 225 506 – 516 [5] Bourke M., et al. (2008) Geomorph. 94 247–255 [6] Thomas P., et al. (2005) Icarus 174 535 – 559 [7] Diniega S., et al. (2009) AGU Fall Meeting Abstracts [8] HiRISE Operations Center Gullies in palikir crater https://www.uahirise.org/dtm/dtm.php?ID=PSP_005943_1380 accessed: 2017-07-03 [11] Byrne, G. F., et al. (1980) Remote Sensing of Env. (10) 175–184 [12] Ingebritsen, S. E. & Lyon, R. J. P. (1985) Int. J. of Remote Sensing (6) 687–696 [13] Rokni, K., et al. (2014) Remote Sensing (6) 4173–4189 [14] Singh, A. (1989) Int. J. of Remote Sensing (10) 989–1003 [15] Sidiropoulos, P. & Muller, J.-P. (2015) PSS (117) 207-222. [16] Parsons, R.A. & Miyamoto, H. (2018) J. of Physics: Conf. Series 1036 (012004).

Additional Information: This study was previously published in Journal of Physics [16].

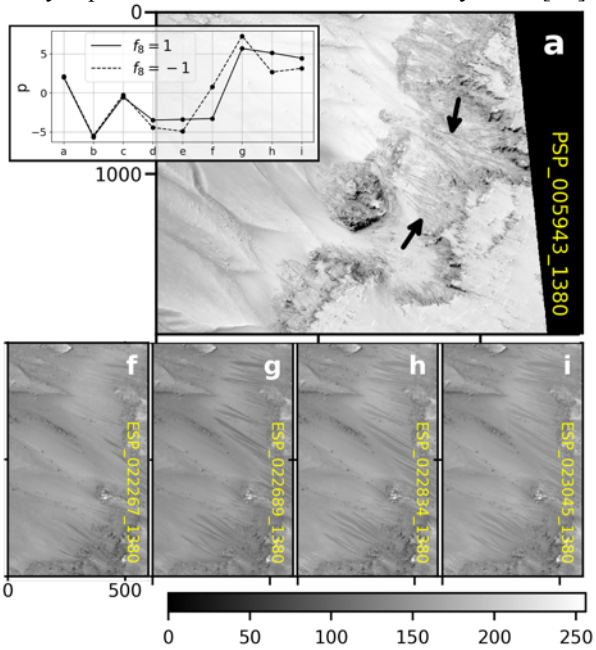

Fig. 3. Cropped sections of images containing RSL (from Fig. 1). Inset gives "feature probability" value (p) for all images.