# Giant Kelp Canopy Biomass from the Landsat Satellite Sensors (TM, ETM+, OLI) Santa Barbara Coastal LTER 2017
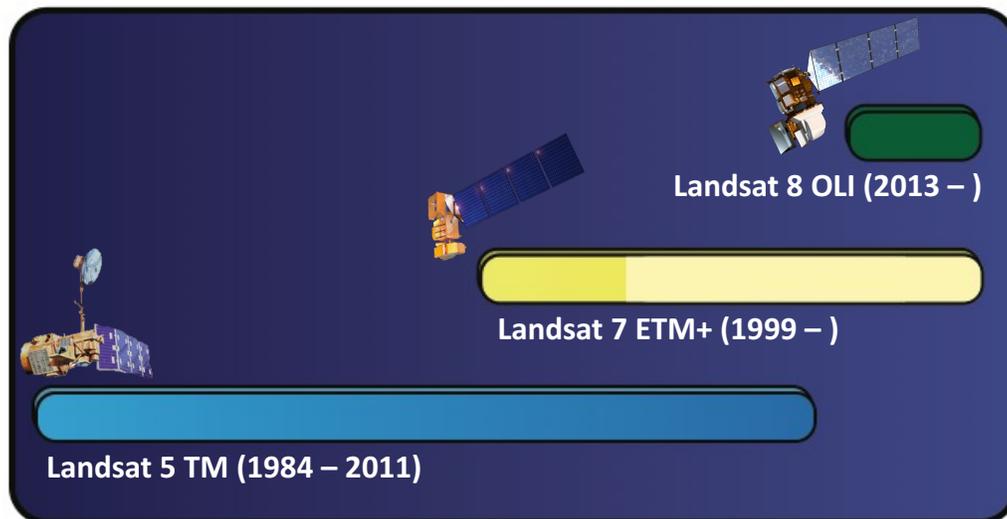


**Figure 1.** Timeline of the three Landsat sensors used in the giant kelp canopy time series. Landsat 5 Thematic Mapper acquired imagery from 1984 – 2011, while Landsat 7 Enhanced Thematic Mapper+ and Landsat 8 Operational Land Imager are both currently in operation. Landsat 7 ETM+ experienced a scan line corrector error in May of 2003, shown as the light-yellow section of the timeline.

*Overview*

The Landsat sensors have acquired 30 m spatial resolution multispectral imagery nearly continuously from 1984 – present, with each sensor imaging the globe every 16 days. We use these images to estimate the canopy biomass of giant kelp (*Macrocystis pyrifera*) along the coast of California, USA. Compared to seawater, emergent kelp canopy presents relatively high reflectance in the near infrared region of the electromagnetic spectrum, and allows for accurate and consistent retrievals of the fraction of kelp canopy in each Landsat pixel across variable ocean conditions. By relating the estimated kelp fraction to long-term diver estimates of canopy biomass, we validated these fractional estimates and applied this relationship to giant kelp forests along the California coast.

*Methods*

The following describes the semi-automated classification process that was developed to estimate giant kelp canopy biomass from Landsat imagery. Images with clear coastline areas were identified and downloaded from the USGS EarthExplorer webpage (earthexplorer.usgs.gov). A single, cloud-free image was selected from the middle of the time series and radiometrically and atmospherically corrected to apparent surface reflectance using the Atmospheric Correction Now (ACORN) software. We standardized the radiometric signals from all other images to this corrected reference image using 50 pseudo-invariant targets that were assumed to be stable across the time series (i.e. airport runways, highways, sandy beaches,

lakes; Furby & Campbell 2001; Baugh & Groeneveld 2008). The number of pseudo-invariant targets was increased to 85 for the Landsat 7 ETM+ images with the SLC corrector error due to missing data lines across the image. Outliers were manually removed to reduce the effects of temporal changes in some of the targets. This procedure accounted for all atmospheric, radiometric, and processing differences between the scenes and created a time series of standardized imagery.

We estimated relative kelp canopy density from the calibrated reflectance data using multiple endmember spectral mixing analysis (MESMA; Roberts et al. 1998). Spectral mixture analysis models the fractional cover of two or more endmembers within a pixel. Each endmember represents a pure cover type and endmembers are assumed to combine linearly (Adams et al. 1993). Standard spectral mixture analysis uses a uniform set of endmembers for the entire image. This approach was problematic for the estimation of giant kelp canopy because of varying water conditions across the image and through time. The reflectance of seawater in the near-shore marine is influenced by sun glint, breaking surface waves, phytoplankton blooms, dissolved organic matter, and suspected sediment. Due to the highly variable seawater reflectance, a single seawater endmember cannot be used.
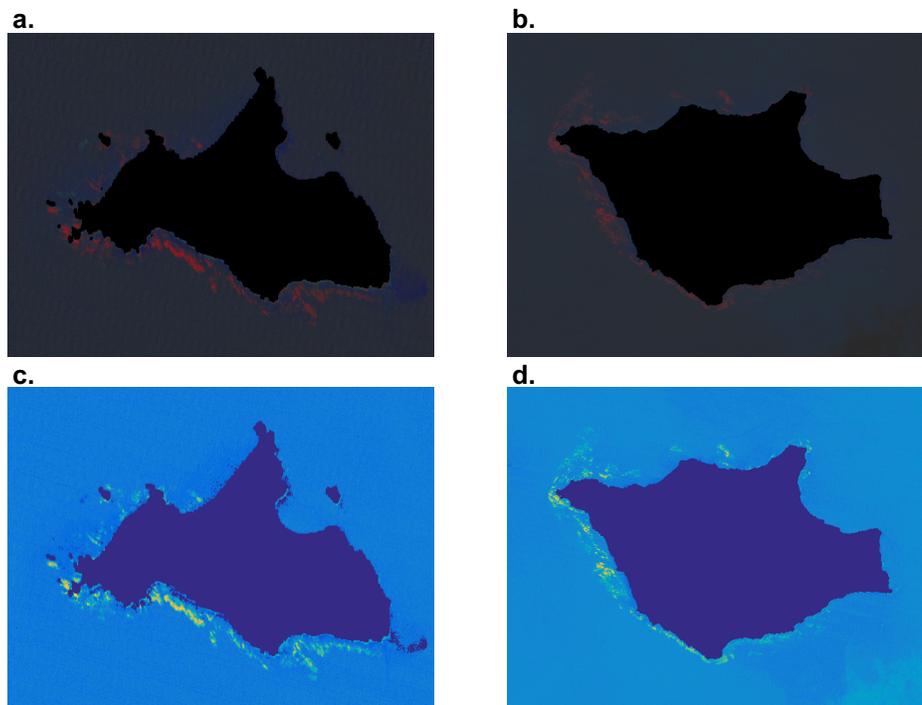


**Figure 2.** Panels a. and b. are false color images of San Miguel and Santa Rosa islands, located in the Santa Barbara Channel. In these images, the near infrared reflectance is displayed as red and floating giant kelp canopy is shown as the orange features along the coasts of the islands. Panels c. and d. show the MESMA kelp fractions derived from the imagery with warmer colors signifying greater fractional cover of kelp canopy in those pixels.

The MESMA process allows endmembers to vary on a per pixel basis by selecting from multiple endmembers for one or more cover types. This technique can better capture the

spectral variability of a cover type through space and time. We modeled pixel reflectance as the linear mixture of reflectance from two endmembers: giant kelp canopy and water. Thirty seawater endmembers were selected from consistently non-kelp covered areas within each Landsat scene. The locations where these endmembers were collected did not change from image date to image date, but the spectral information collected at these locations varied between images. A single kelp endmember was selected by extracting kelp covered pixel spectra from each image and finding the single spectrum that fit the entire library of kelp spectra with the lowest root mean square error (RMSE; Dennison & Roberts 2003). The pixels of each image were modeled as a two-endmember mixture of kelp and each of the 30 water endmembers that were free of cloud contamination. The final model (out of 30) chosen for each pixel was the model that minimized the RMSE when fit to the spectrum of that pixel. The result of this process was a measure of the relative fraction of each pixel covered by kelp canopy (Figure 2). The MESMA process successfully estimated the relative canopy fraction of giant kelp under a variety of conditions including large amounts of sediment runoff and high levels of sun glint (Cavanaugh et al. 2011).
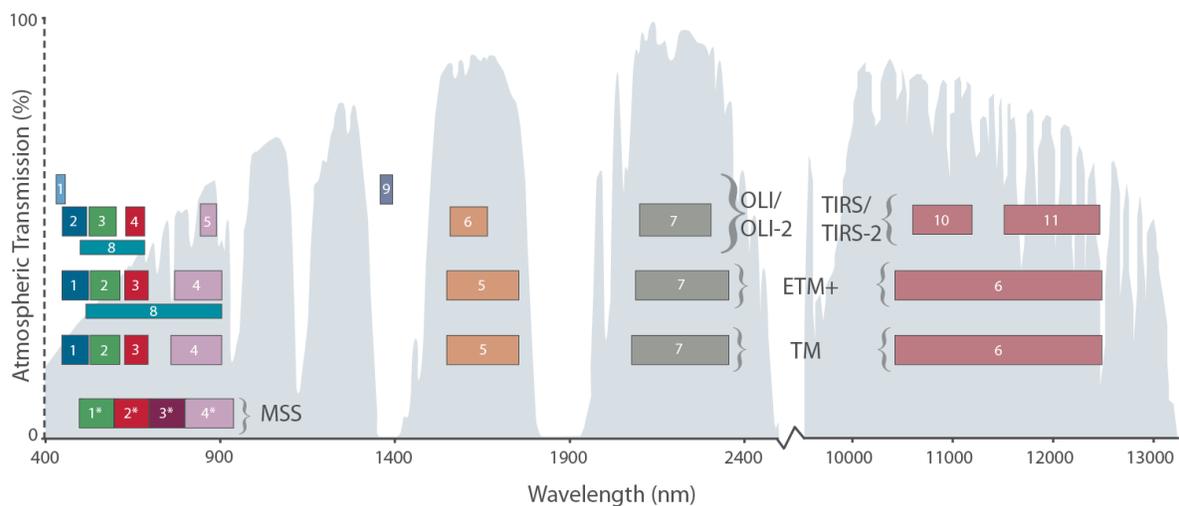


**Figure 3.** Image showing the spectral bands of the three sensors used in this dataset. Note the spectral band width in the Landsat 8 OLI sensor compared to Landsat 5 TM and Landsat 7 ETM+. This difference is most apparent in the near infrared band 5 for OLI, compared to band 4 for TM and ETM+. Bands 1 – 4 are used for the MESMA process for TM and ETM+ and bands 2 – 5 are used for OLI.

Each Landsat sensor differs in signal to noise ratio, radiometric calibration, and the number and width of their spectral bands. The most important difference for the Landsat kelp time series was the "optimization" of the Landsat 8 OLI spectral bands relative to the Landsat 5 TM and Landsat ETM+ sensors (Figure 3). Due to the lack of temporal overlap between Landsat 5 and Landsat 8, the dynamic nature of the giant kelp canopy, and the fact that each image is acquired 8 days apart for the overlapping sensors, the best way to compare the kelp fraction retrievals from each sensor was by using simulated imagery. We took advantage of aerial hyperspectral imagery over the Santa Barbara Channel collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) mounted on an ER-2 aircraft flying at 65,000 feet to simulate

hyperspectral satellite imagery. We resampled a hyperspectral image collected in April 2013 to all three Landsat sensors' spectral bands using published spectral response functions. The MESMA process was then applied to each simulated image using 30 seawater endmembers from the same locations in each simulated image and the kelp endmember discussed above. Kelp fractions were then compared between each simulated sensor pair (Figure 4). Kelp fractions estimated from the Landsat 5 TM and Landsat 7 ETM+ were comparable however Landsat 8 OLI kelp fractions were consistently lower for the same amount of kelp canopy. To address this difference, we adjusted the Landsat 8 kelp fractions using equation (1) to match those of the previous Landsat sensors.
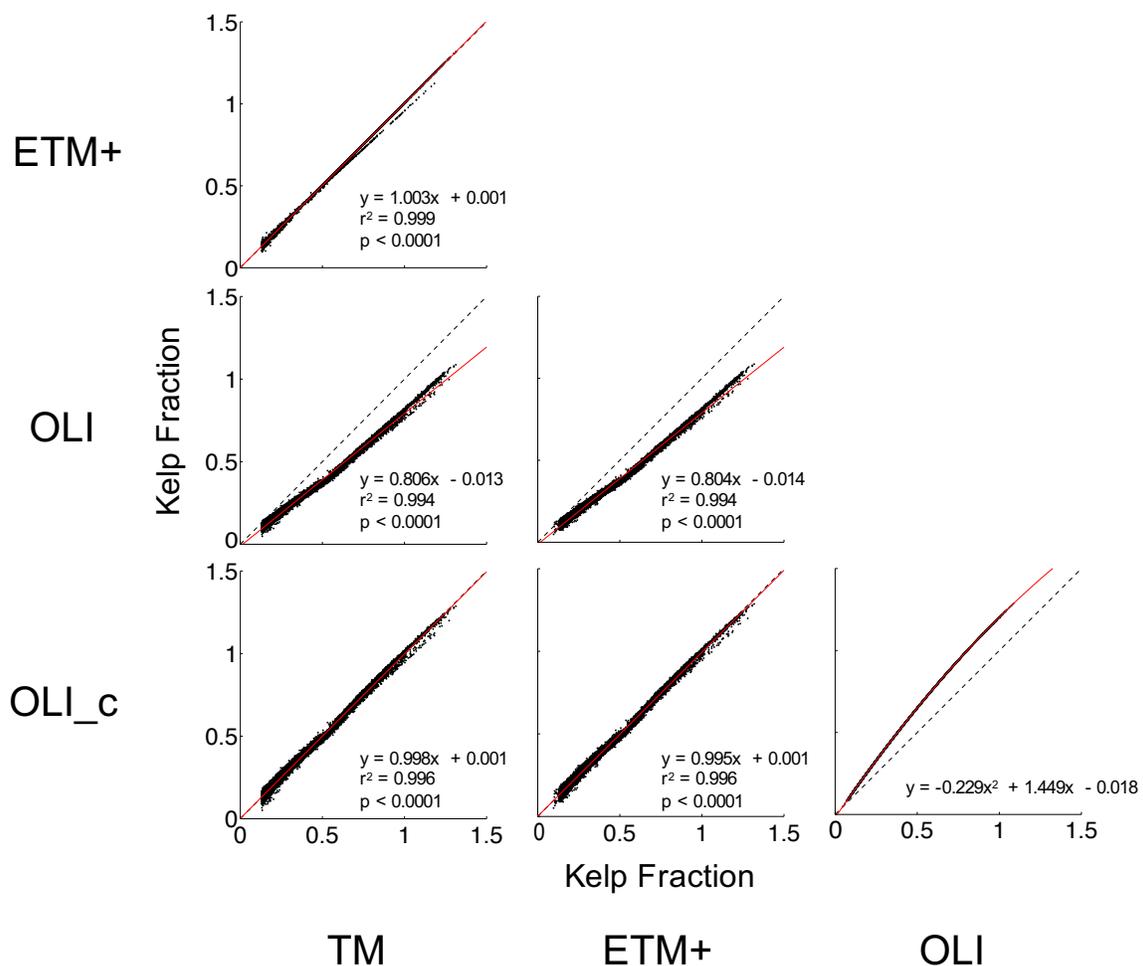
$$y = -0.229x^2 + 1.449x - 0.018$$

Eq. 1



**Figure 4.** Scatterplot matrix of MESMA derived kelp fractions from the simulated images of each Landsat sensor, compared against every other Landsat sensor used in the kelp canopy biomass time series. The dashed black lines show the 1:1 line while the red lines are the best fit lines for each individual scatterplot. TM represents the Landsat 5 Thematic Mapper, ETM+ is the Landsat 7 Enhanced Thematic Mapper +, OLI is the Landsat 8 Operational Land Imager, and OLI_c it the Operational Land Imager kelp fraction corrected to match the TM and ETM+ kelp fraction estimates.

The retrieved kelp fractions from each sensor were then compared to giant kelp canopy biomass observations that were collected by divers at permanent plots maintained by the Santa Barbara Coastal Long Term Ecological Research (SBC LTER) project at the Mohawk and Arroyo Quemado kelp forests (Figure 5). The data and the methods used to measure giant kelp canopy biomass from diver surveys are described in detail in Rassweiler et al. (2008). Briefly, divers measured the length of all fronds along 5 transects (40 x 1 m) within a plot (40 x 40 m) and converted these lengths to biomass using validated length to weight relationships. Each plot was overlapped by four 30 m Landsat pixels. For each image, we compared the diver measured canopy biomass of each plot to the mean kelp fraction of the four pixels, weighted by the proportion of the transect area in each pixel. We compared biomass estimates to kelp fractions from each sensor, and all sensor measurements together, using reduced major axis linear regressions (Figure 6).
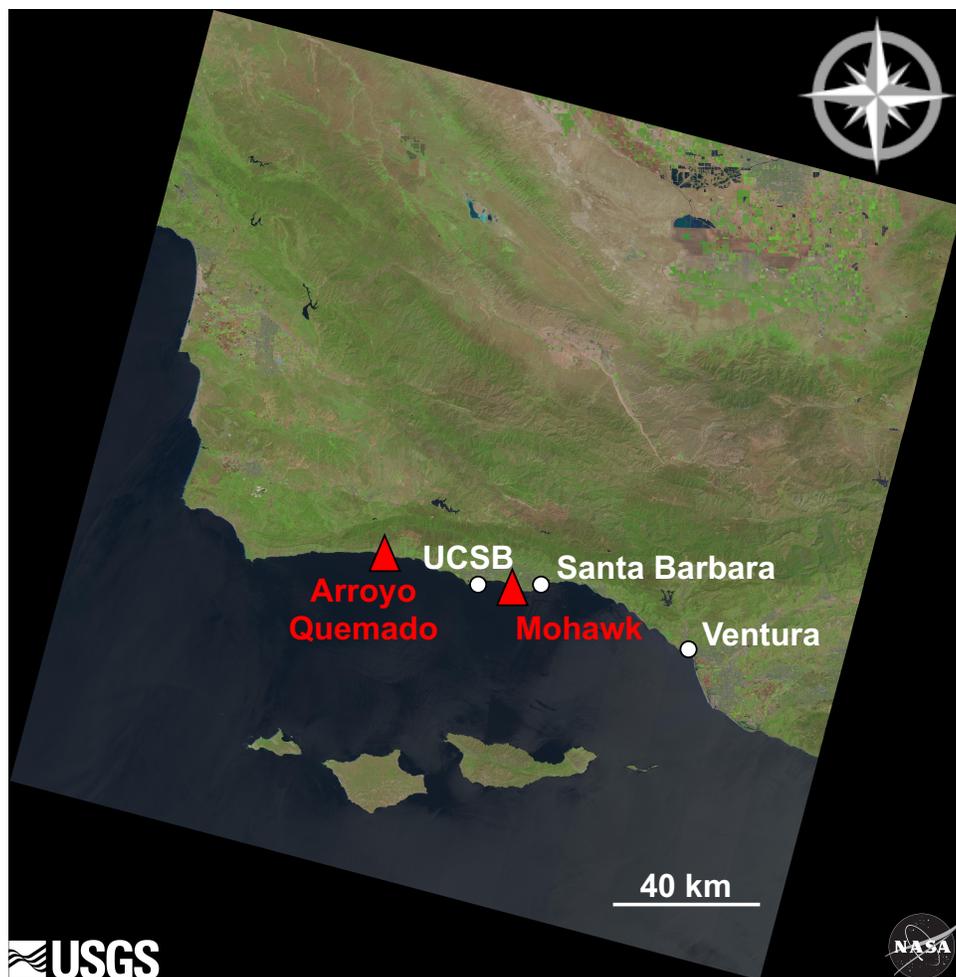


**Figure 5.** Landsat 8 Operational Land Imager image displaying a portion of the study area, including the two SBC LTER research sites (red triangles) where diver surveys of canopy biomass were conducted.
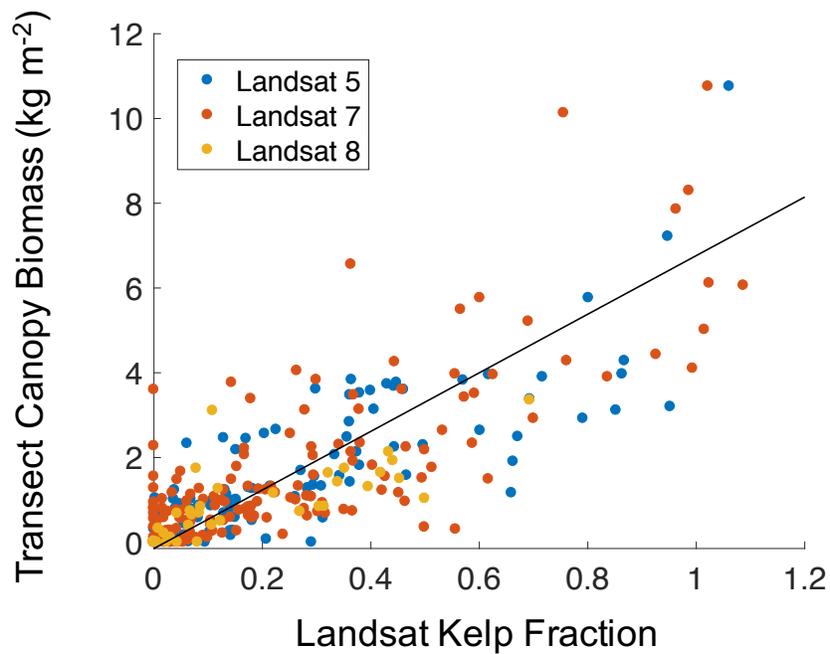
**Figure 6.** Validation of Landsat satellite estimates of the three sensors versus diver estimated canopy biomass from the two SBC LTER sites. The black line represents the reduced major axis linear regression line across all three sensors.

A strong positive linear relationship was found between the Landsat derived kelp fractions and giant kelp canopy biomass across all three sensors (Table 1). The relationship between kelp fraction and canopy biomass using the Landsat 8 sensor was significantly lower than that relationships using the other two sensors. However, since the Landsat 8 satellite began to acquire imagery in mid-2013, the SBC LTER calibration sites have not recorded canopy densities higher than 3.5 kg m$^{-2}$. When the relationships were examined between 0 – 3.5 kg m$^{-2}$, all three sensors presented equations that were statistically similar. This result hints at a nonlinear relationship between kelp fraction and canopy biomass which will be examined upon the next dataset update. Since all relationships were similar across the range of canopy biomass measured by the three sensors we were confident in using a common equation (Figure 6), which we used to transform images of kelp fractional cover into quantitative, validated maps of giant kelp canopy biomass.

| | Landsat 5 TM | Landsat 7 ETM+ | Landsat 8 OLI | All |
|---|---|---|---|---|
| **Canopy Biomass** $(r^2)$ | 0.67 | 0.63 | 0.57 | 0.64 |
| **Equation (All Canopy Biomass)** | y = 6.54x - 0.09 (0.37, 0.13) | y = 7.30x - 0.17 (0.37, 0.13) | y = 4.68x + 0.05 (0.54, 0.13) | y = 6.91x - 0.15 (0.25, 0.09) |
| **Equation ($<$3.5 kg m$^{-2}$)** | y = 4.64x + 0.16 (0.34, 0.10) | y = 5.01x + 0.09 (0.37, 0.09) | y = 4.68x + 0.05 (0.54, 0.13) | |

**Table 1.** The coefficient of determination and reduced major axis linear regression line equations for each Landsat sensor and across all sensors. All relationships are significant at the $p < 0.001$ level.

Upon examination of the regional kelp biomass data, inconsistencies in biomass estimates were apparent between the Landsat 5 and 7 sensors. These inconsistences were not stable across space with some Landsat scenes showing higher biomass estimates for one sensor, while a nearby scene would show the opposite. We were only able to fully compare Landsat 5 and 7 canopy biomass estimates between 1999 – 2003 (see *ETM+ SLC Error Gap Filling* below). These discrepancies were attributed to the 8-day repeat difference between the satellites synching with tidal cycles across this time range (Figure 7). Pixel-based canopy biomass estimates were aggregated into coastline segments and compared inside the San Diego, Los Angeles, and Santa Barbara Landsat scenes. Each Landsat scene showed a biomass difference consistent in sign and magnitude with a tidal effect, and when each segment biomass was adjusted based on local tidal patterns, these effects were alleviated (Figure 8).
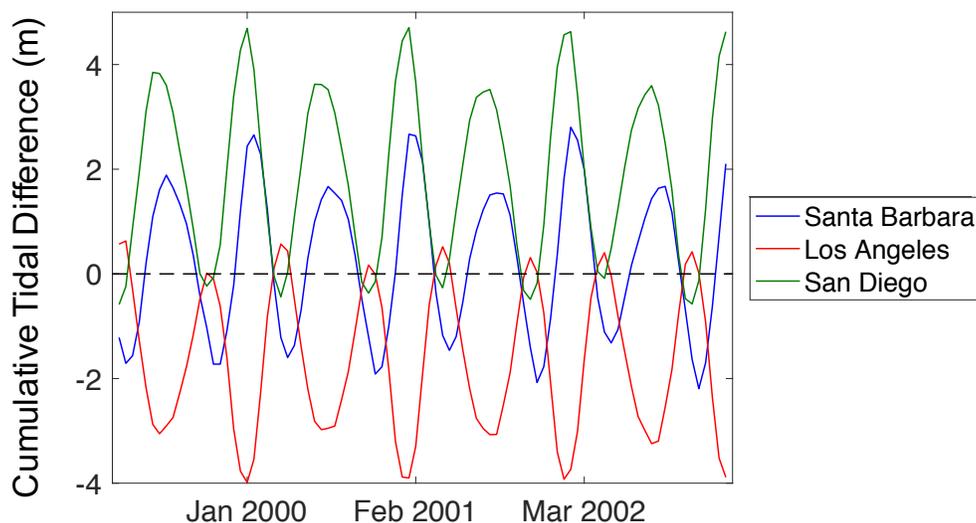


**Figure 7.** The cumulative difference in tidal height between Landsat 5 and Landsat 7 sensors during the overlap period without the scan line corrector error. The dashed black line shows the zero line where the cumulative difference between the sensors is zero. If the lines are over the zero line Landsat 7 was at a higher tidal state than Landsat 5 and if the lines are below the zero line, the opposite is true.
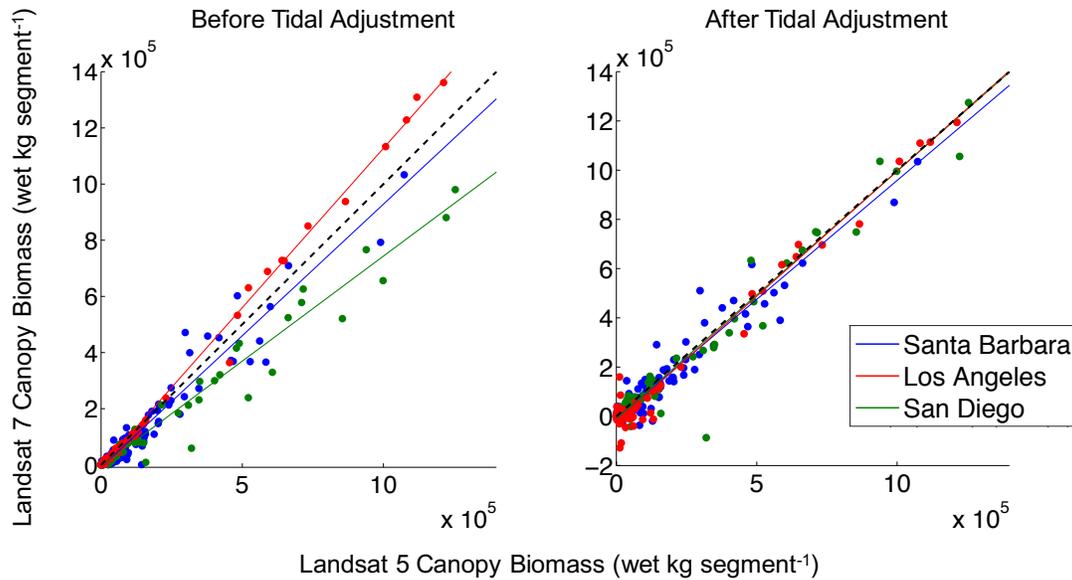
**Figure 8.** Relationships between the summed canopy biomass of all 500m coastline segments across three Landsat scenes, both before and after tidal height adjustment. The dashed black line is the 1:1 line and the colored lines are best fit regression lines.

Cloud-free imagery of the California coastline is available about every 1-2 months during the time series. This frequency increases when multiple satellites are acquiring imagery. Rather than correct for tidal state explicitly, we decided to take the mean of all biomass estimates within a quarter (3-months). The current version of the dataset includes eight Landsat scenes, which cover the entire region of dominance for giant kelp in California, roughly from Año Nuevo Island to the US/Mexico border (Figure 9). The dataset is provided as a netCDF file that includes the mean canopy biomass of every identified kelp containing pixel, for each quarter, from 1984 – 2015, along with relevant metadata.

**Figure 9.** Landsat 8 OLI mosaic showing the coastal area covered by this dataset, from Año Nuevo Island in the north, to the US/Mexico border.

*ETM+ SLC Error Gap Filling*

In mid-2003 the scan line corrector (SLC) of the Landsat 7 ETM+ sensor failed. This corrector compensates for the forward motion of the satellite and merges the imagery collected from the 'whisk broom' type sensor into a complete image. With the permanent failure of this hardware, a zig-zag pattern of missing data lines is present in each image, with the width of the

lines increasing towards the image edge (Figure 10). The location of the missing data lines change from image date to image date.
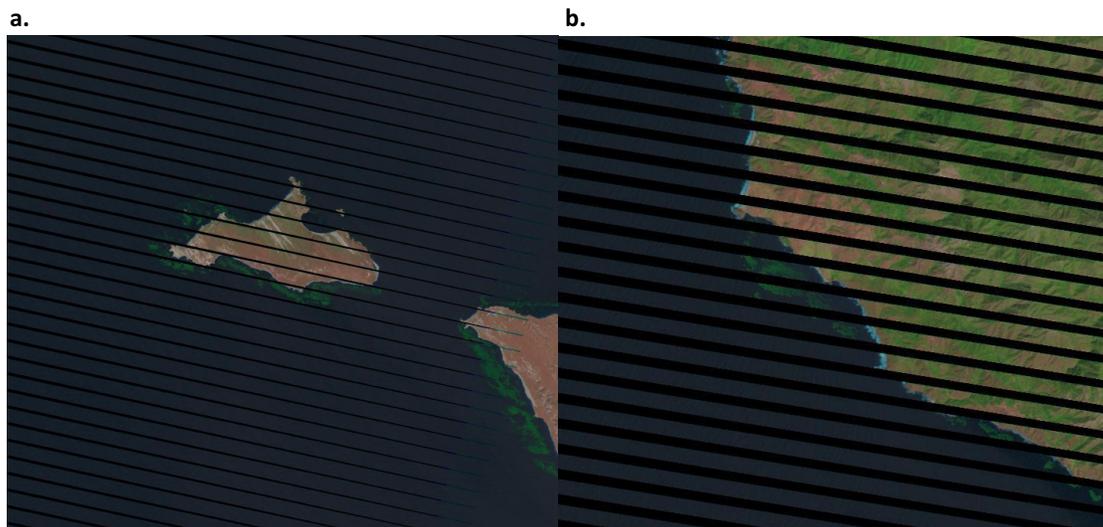
a.                          b.



**Figure 10.** a. Scan line corrector error lines near the center of the Landsat image tile showing missing data lines over San Miguel and Santa Rosa islands in the Santa Barbara Channel. b. Error lines near the edge of a Landsat image tile, showing thicker missing data lines near Point Sur in central California. Giant kelp canopy is shown as the bright green features along the coast in both images.

Soon after the failure of the SLC, the United States Geological Survey implemented a Phase 1 gap filling methodology using a localized linear histogram technique to fill the scan line gap with previously acquired Landsat 7 imagery (Scaramuzza et al. 2004). Since the maximum width of a scan line gap is 14 pixels, a 17-pixel wide window was chosen to move across the missing pixels and use the valid pixels inside the window to establish a corrective gain and bias value for the missing pixel in the center on the window. This corrective gain and bias was then applied to valid co-registered pixels from a Landsat 7 image captured close in time to fill the missing pixels with corrected spectral data. Soon after this Phase 1 methodology was implemented, it was replaced with a Phase 2 gap filling algorithm using an adaptive window logic for increased processing speed as well as allowing users to choose multiple images as valid fill data (Storey et al. 2005). Heterogeneous landscapes necessitated the development of a segmentation model that divided landscapes into contiguous sub-regions and avoided the use of square windows for averaging (Maxwell et al. 2007). Segmentation models were developed from global cover maps and missing SLC pixels were filled with the dominant spectral values from inside each of three progressively larger segments. Later, a similar pixel method was employed that searched for a missing pixel in a Landsat 5 TM image close in time. Pixels which resembled the targeted pixel were identified in the Landsat 5 image within a 5 x 5 pixel window, which increased in size if these similar pixels could not be found in the Landsat 7 image. Similar pixels were weighted based on similarity in spectral shape and distance to the target pixel and were then used to predict the spectral information for the missing pixel (Chen et al. 2011). Following this progression in gap filling methods, we aimed to estimate giant kelp biomass in missing data pixels using known relationships through both space and time.

Giant kelp canopy biomass is known to display high, but exponentially decreasing, spatial synchrony over the first several hundred meters in distance (Figure 11; Cavanaugh et al. 2013). We leveraged this phenomenon to predict canopy biomass in missing pixels using a combination of the biomass state of nearby pixels and the relationship between these nearby pixels and the missing pixel through time. We first determined all known kelp pixels that were covered by a scan line gap across the Landsat time series. A known kelp pixel is defined as any pixel that has contained kelp canopy at least five times in the combined Landsat 5 and 7 time series. All pixels within a 300-meter radius from the missing pixel are located and the linear relationship of biomass through time between the missing pixel and each nearby pixel is found using a reduced major axis linear regression. For each significant relationship with a correlation coefficient > 0.8, a biomass estimate for the missing pixel is generated using the regression slope and offset. The mean of these estimates is used as the missing pixel fill value and the standard error is retained as a measure of uncertainty. Missing pixels where > 70% of nearby pixels show zero detected canopy biomass are filled with a value of zero. Missing pixels with no nearby kelp pixels with correlation coefficients > 0.8 are filled using a piecewise cubic interpolation of the missing pixel through time.
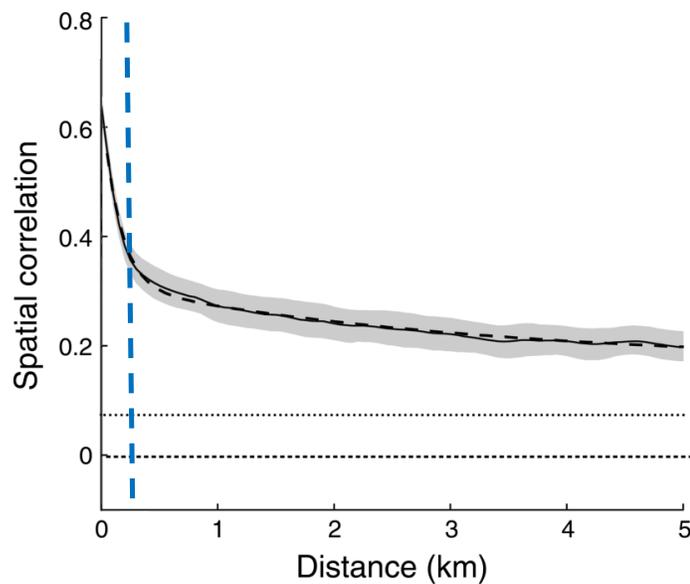


**Figure 11.** Nonparameric spatial correlation function (solid line) and double exponential fit (dashed black line) for giant kelp canopy biomass (adapted from Cavanaugh et al. 2013). Grey shaded area shows the 95% bootstrapped confidence interval. The blue dashed line shows the 300 m cutoff for the spatial synchrony-based gap filling algorithm.

To validate this synchrony-based gap filling algorithm we selected six Landsat 5 images across the study period and masked out pixels using a scan line gap mask from a Landsat 7 image (Figure 12). The dates used for validation were: November 16th, 2000; October 5th, 2002; July 14th, 2004; November 22nd, 2005, August 8th, 2008; and January

17<sup>th</sup>, 2009. We then filled these missing pixels and validated the predicted biomass using the actual canopy biomass measured by the sensor.
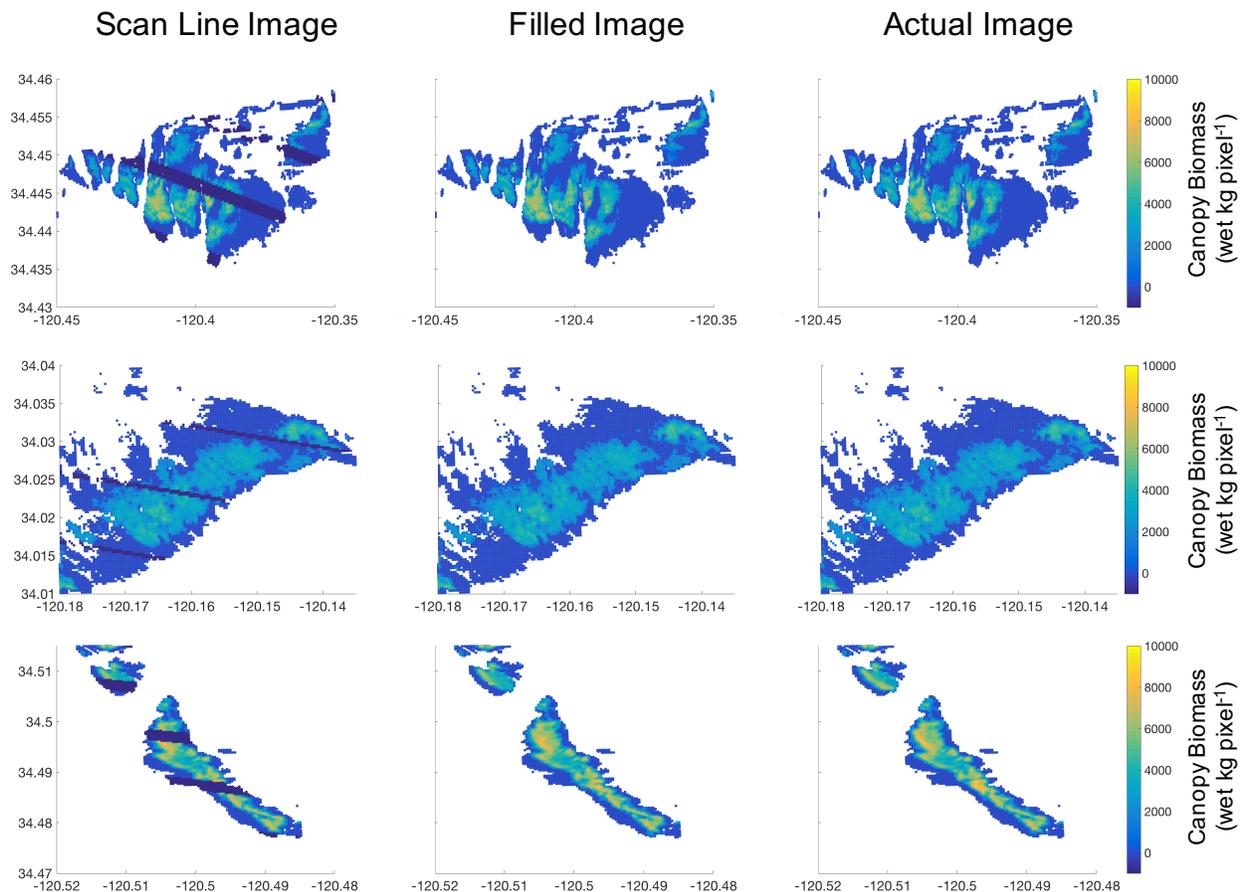


**Figure 12.** Results of gap filling algorithm across three kelp forest canopies in the Landsat 5 validation images. In the first column, kelp forest pixels are masked with multiple dark blue simulated scan line error missing data lines. The second column shows the results of the gap filling algorithm. The third column shows the actual images before the scan line error missing data lines.

Across the six dates, the algorithm was used to fill 80,286 missing kelp pixels. Of that total, 59% were filled using spatial synchrony and 41% were filled using temporal interpolation based on whether the pixel to be filled had nearby kelp pixels with high spatial synchrony. Pixels filled using spatial synchrony had a correlation coefficient of 0.91 ($p < 0.0001$) and a linear equation of $y = 0.94x + 74$ (Figure 13). Pixels filled using temporal interpolation had a correlation coefficient of 0.73 ($p < 0.0001$) and a linear equation of $y = 0.74x - 6$. The overall relationship had a correlation coefficient of 0.88 ($p < 0.0001$) and a linear equation of $y = 0.92x + 16$. The results show that the pixels filled using the spatial synchrony method were closer to actual canopy biomass than those filled using interpolation between two close dates. Overall, the total gap filling algorithm performed well in estimating the canopy biomass on a pixel scale, leading to general confidence in the algorithm to fill the scan line missing data gaps.
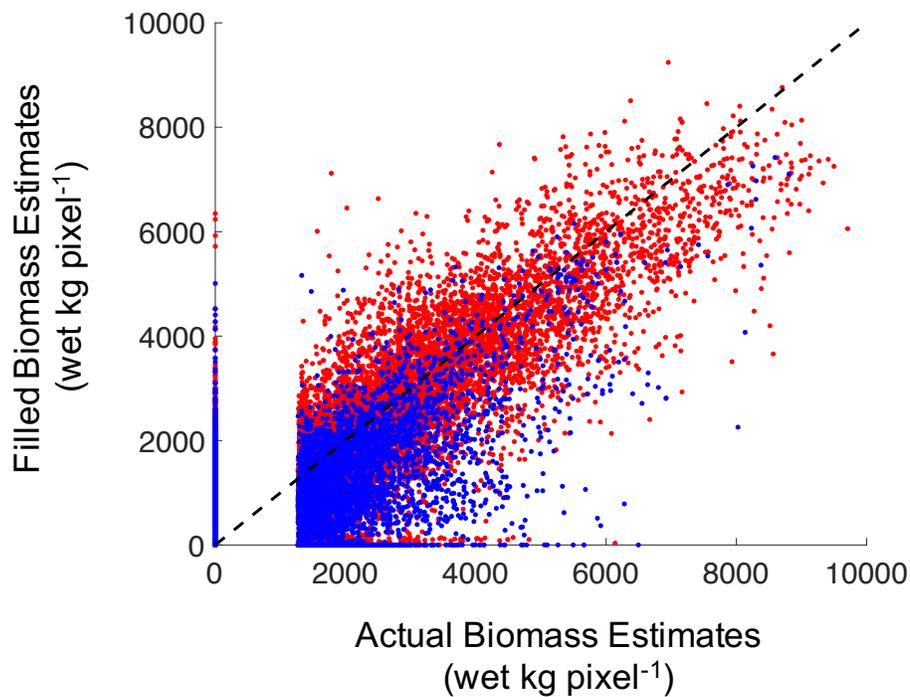
**Figure 13.** Actual versus filled pixel values from the validation imagery. Red points show pixels filled using spatial synchrony and blue points show pixels filled using temporal interpolation. The dashed line is the 1:1 line. The section of no data between 0 and 1300 kg on the x-axis represents the minimum detection limit for the Landsat satellites.

Since most studies have combined pixels together into coastline segments or patches, we also compared the total canopy biomass of coastline segments affected by scan line missing data gaps. We aggregated pixels into 500-meter coastline segments by assigning each pixel to its closest coastline point along a 500-meter grid. We then determined if part of that segment contained any missing data lines and excluded those that did not from the analysis (Figure 14). The overall relationship had a correlation coefficient of 0.98 (p < 0.0001) and a linear equation of y = 0.98x + 16000. This analysis shows that the gap filling algorithm provides excellent data for studies that aggregate data into coastline segments or patches. One cautionary note is that the algorithm did assign biomass to some coastline segments which have zero biomass in the actual imagery. This is probably due to fill values generated from kelp pixels in adjoining segments.
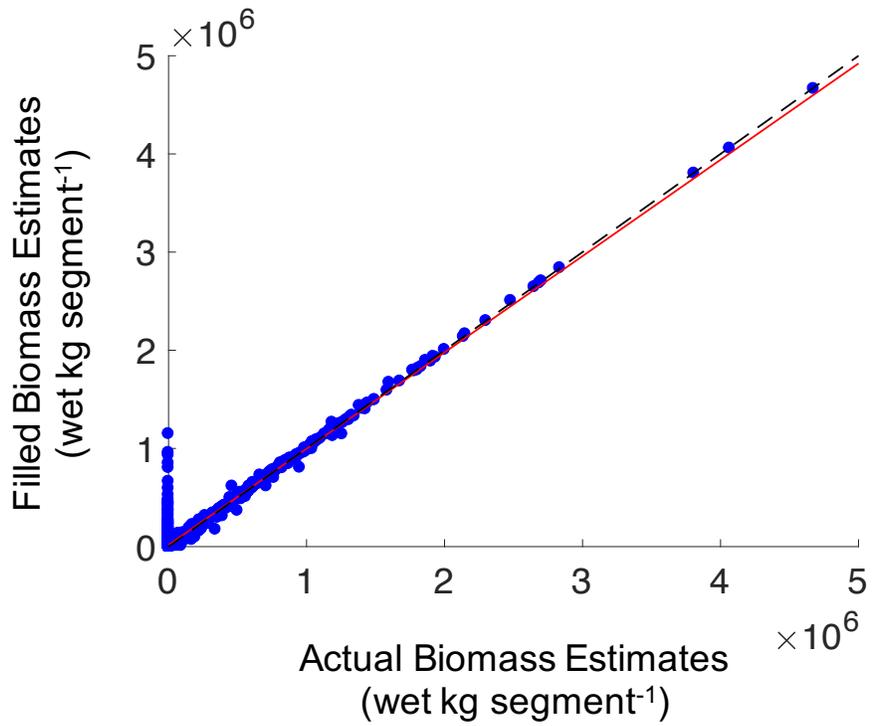
**Figure 14.** Actual versus filled coastline segment values from the validation imagery. The dashed line is the 1:1 line and the red line is the best fit line.

*References*

Adams, J., Smith, M., and A. Gillespie. 1993. Imaging spectroscopy: Interpretation based on spectral mixture analysis. Remote Geochemical Analysis: Elemental and mineralogical composition. **7**: 145–166.

Baugh, W., and D. Groeneveld. 2008. Empirical proof of the empirical line. Int. J. Remote Sens. **29**: 665–672.

Cavanaugh, K. C., B. E. Kendall, D. A. Siegel, D. C. Reed, F. Alberto, and J. Assis. 2013. Synchrony in dynamics of giant kelp forests is driven by both local recruitment and regional environmental controls. Ecology **94**: 499–509. doi:10.1890/12-0268.1

Cavanaugh, K., D. Siegel, D. Reed, and P. Dennison. 2011. Environmental controls of giant-kelp biomass in the Santa Barbara Channel, California. Mar. Ecol. Prog. Ser. **429**: 1–17. doi:10.3354/meps09141

Chen, J., X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin. 2011. Remote Sensing of Environment A simple and effective method for fi lling gaps in Landsat ETM + SLC-off images. Remote Sens. Environ. **115**: 1053–1064. doi:10.1016/j.rse.2010.12.010

Dennison, P., and D. Roberts. 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sens. Environ. **87**: 123–135.

Furby, S. L., and N. A. Campbell. 2001. Calibrating images from different dates to "like-value" digital counts. Remote Sens. Environ. **77**: 186–196. doi:10.1016/S0034-4257(01)00205-X

Maxwell, S. K., G. L. Schmidt, and J. C. Storey. 2007. A multi-scale segmentation approach to filling gaps in Landsat ETM + SLC-off images. Int. J. Remote Sens. **28**: 5339–5356. doi:10.1080/01431160601034902

Rassweiler, A., K. K. Arkema, and D. C. Reed. 2008. Net Primary Production, Growth, and Standing Crop of Macrocystis pyrifera in Southern California. Ecology **89**: 2068–2068.

Roberts, D., M. Gardner, and R. Church. 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. Remote Sens. … **65**: 267–279.

Scaramuzza, P., E. Micijevic, and G. Chander. 2004. SLC Gap-Filled Products Phase One Methodology. USGS 1–5.

Storey, J., P. Scaramuzza, and G. Schmidt. 2005. Landsat 7 Scan Line Corrector-Off Gap-Filled Product Development. Pecora **16**: 23–27.