ISPRS Journal of Photogrammetry and Remote Sensing 98 (2014) 106–118

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

An effective approach for gap-filling continental scale remotely sensed time-series

Daniel J. Weiss [a,*], Peter M. Atkinson [b], Samir Bhatt [a], Bonnie Mappin [a], Simon I. Hay [a,c], Peter W. Gething [a]

[a] Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, Oxford, UK
[b] Geography and Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK
[c] Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form 9 October 2014
Accepted 9 October 2014
Available online 6 November 2014

Keywords:
Gap-filling
MODIS
EVI
LST
Africa

ABSTRACT

The archives of imagery and modeled data products derived from remote sensing programs with high temporal resolution provide powerful resources for characterizing inter- and intra-annual environmental dynamics. The impressive depth of available time-series from such missions (e.g., MODIS and AVHRR) affords new opportunities for improving data usability by leveraging spatial and temporal information inherent to longitudinal geospatial datasets. In this research we develop an approach for filling gaps in imagery time-series that result primarily from cloud cover, which is particularly problematic in forested equatorial regions. Our approach consists of two, complementary gap-filling algorithms and a variety of run-time options that allow users to balance competing demands of model accuracy and processing time. We applied the gap-filling methodology to MODIS Enhanced Vegetation Index (EVI) and daytime and nighttime Land Surface Temperature (LST) datasets for the African continent for 2000–2012, with a 1 km spatial resolution, and an 8-day temporal resolution. We validated the method by introducing and filling artificial gaps, and then comparing the original data with model predictions. Our approach achieved R^2 values above 0.87 even for pixels within 500 km wide introduced gaps. Furthermore, the structure of our approach allows estimation of the error associated with each gap-filled pixel based on the distance to the non-gap pixels used to model its fill value, thus providing a mechanism for including uncertainty associated with the gap-filling process in downstream applications of the resulting datasets.

1. Introduction

Past and current data collection efforts have produced numerous remotely sensed imagery time-series, often exceeding a decade in length, with tremendous utility (both realized and potential) for a wide range of research applications (Hay et al., 2006; Scharlemann et al., 2008). However, gaps within such time-series reduce the utility of these data sources for modeling and monitoring environmental phenomena, and gaps are particularly problematic within imagery of tropical and sub-tropical areas where persistent cloud-cover can obscure portions of the landscape seasonally or throughout the year. Gaps within fine temporal resolution time-series such as those derived from NASA's Moderate Resolution Imaging Spectrometer (MODIS) imagery have been partially filled through the creation of products that summarize daily data into multi-day composites (e.g., 8- or 16-day). However, in the cloudiest of areas even composite products often contain problematic gaps, and these gaps take on added significance as they tend to occur in areas (e.g., equatorial Africa or the Amazon basin) for which few alternative geospatial datasets exist for characterizing dynamic landscape processes.

Our goals in this research were to develop a data-driven gap-filling methodology that (1) balances the need for accuracy with the computational efficiency necessary for feasible application to continental-scale time-series, (2) uses both spatial and temporal information within the data time-series to fill the gap pixels, (3) requires no ancillary datasets such as land cover products or digital elevation models to model missing pixel values, and (4) provides a standardized yet flexible approach that is applicable to a wide range of datasets. Among these goals, the first was most relevant to the wider remote sensing community as the large data volume associated with continental-scale time-series limits the utility of mathematically complex (e.g., geostatistical) algorithms for rapid gap-filling. Expected ancillary benefits of a conceptually simple approach include increased accessibility to a wider audience of

* Corresponding author.
E-mail address: daniel.weiss@zoo.ox.ac.uk (D.J. Weiss).

http://dx.doi.org/10.1016/j.isprsjprs.2014.10.001
0924-2716/

potential image time-series users, as well as ease of adaptation of the developed methods for use with new datasets.

The gap-filling approach ultimately developed in this research is predicated on using both neighboring (non-gap) data and data from other time periods (i.e., calendar date or multi-year summary datasets) to fill gaps within image time-series. Our underlying hypothesis was that spatial and temporal autocorrelation inherent within longitudinal imagery archives can be leveraged to gap-fill remotely sensed data products. We developed and tested the gap-filling methodology using the MODIS Enhanced Vegetation Index (EVI) and Land Surface Temperature (LST) 1 km products, acquired for the African continent, from 2000–2012, with an 8-day temporal resolution. These data products were selected for eventual use in modeling malaria risk in Africa, but they are potentially useful for many research endeavors given their widespread utility. In particular, LST is correlated with air temperature (Mildrexler et al., 2011) and EVI is useful as a proxy (albeit lagged in time) for moisture in Africa (Jamali et al., 2011). Africa was selected as our study area because substantial portions of the continent experience widespread seasonal cloud cover, making this both an ideal region to test the methodology and an area in need of gap-filled products. Furthermore, processing time-series data for the whole of Africa presents a rigorous computational test for the presented gap-filling method.

2. Background

Numerous gap-filling approaches have been developed for modeling erroneous or missing data caused by clouds, shadows, or sensor malfunctions. These approaches can be roughly divided into the following categories: (1) methods that rely on spatial information, (2) methods based on temporal information available within an image time-series, and (3) methods that include both spatial and temporal information in the gap-filling process. Examples exist within each of these categories that include ancillary information, such as imagery from another sensor, a digital elevation model, or a classified land cover dataset, within the modeling process.

2.1. Spatial gap-filling approaches

Geostatistical approaches such as kriging have long been utilized for gap-filling imagery using the information present within surrounding (non-gap) pixels to interpolate missing data (e.g., Addink, 1999; Rossi et al., 1994). Introducing a second, gap-free dataset (e.g., an image from the same sensor acquired for the area of interest on a different date) enables gap-filling using cokriging techniques (Zhang et al., 2007, 2009) as well as gap-filling approaches predicated on image segmentation (Bédard et al., 2008; Maxwell, 2004; Maxwell et al., 2007). Using data from an alternative date is also the technique underlying the novel Neighborhood Similar Pixel Interpolator method for filling gaps in Landsat ETM+ imagery developed by Chen et al. (2011), which was later augmented to include geostatistical theory by Zhu et al. (2012).

2.2. Temporal gap-filling approaches

The second category of gap-filling approaches relies on modeling missing pixel values using values associated with the missing pixel from different points in time, and a comparison of temporal approaches is provided in an informative review by Kandasamy et al. (2012). Jönsson and Eklundh (2004) made an important contribution to temporal approaches by developing the TIMESAT software package, which contains built-in asymmetric Gaussian and Savitzky–Golay filters for smoothing time-series data as a means of estimating missing data. Notable examples of temporal gap-filling applications include approaches for gap-filling MODIS Leaf Area Index (LAI) data (Gao et al., 2008) and NDVI derived from AVHRR data (Roerink et al., 2000). More recently Verger et al. (2013) developed the Consistent Adjustment of the Climatology to Actual Observations approach for increasing the accuracy of temporal interpolations of missing LAI data derived from AVHRR imagery by including climatological data within the model.

2.3. Spatio-temporal gap-filling approaches

Several spatio-temporal gap-filling approaches have been developed that utilize multi-step modeling approaches whereby the algorithm fills missing values using an alternating sequence of purely spatial or temporal steps. Kang et al. (2005) developed such an approach for gap-filling ecosystem metrics (i.e., fPAR, LAI, and net photosynthesis) modeled from MODIS data using simple spatial interpolation within land cover classes. If no cloud-free pixels were found within a 5 by 5 pixel window, the algorithm used temporal interpolation to fill the pixel using data from earlier and later dates. Borak and Jasinski (2009) later used a modified version of the Kang et al. (2005) approach when gap-filling MODIS LAI for a large portion of North America. Gafurov and Bárdossy (2009) also developed a stepped approach for gap-filling the MODIS snow cover product, but unlike the Kang et al. (2005) approach the algorithm developed by these authors prioritizes temporal gap-filling models and also includes a step that incorporates pixel elevation. More recently Poggio et al. (2012) developed an innovative method for gap-filling MODIS EVI data that utilizes a hybrid Generalized Additive Model (GAM) – geostatistical space–time model to model missing pixel values using spatial (latitude, longitude and elevation) and temporal (date of year) information as model covariates.

3. Materials and methods

From our review of existing gap-filling methodologies we identified the Chen et al. (2011) approach as the most promising starting point for gap-filling the MODIS time-series of Africa due to its relative simplicity and computational efficiency. The immediate challenge in adapting this approach was to develop a fully operational algorithm capable of processing time-series data at a continental scale within a several-month time frame. Given these time constraints and the data volume of the project (i.e., nearly a terabyte in size) we ultimately developed two complementary algorithms that fill gaps by utilizing ratios from neighboring (non-gap) pixels derived at two points in time, similar to Chen et al. (2011), but modified for use with single-banded MODIS time-series to increase processing speed. The approach (Fig. 1 – explained in detail below) we develop (1) ingests raw images, (2) finds gap pixels that may first be identified using a despeckling algorithm, (3) fills some pixels using an algorithm that relies on calendar data imagery, and (4) fills the remaining gap pixels using a second algorithm that runs much faster by leveraging processing already used to fill adjacent gaps. Our gap-filling approach produces three output datasets for each image within a time-series: (1) a gap-filled image, (2) a flag image identifying the algorithm (if any) that was used for each pixel, and (3) a distance image quantifying the spatial lag between the filled pixel and the neighboring pixels used in the gap-filling model. We validated the approach by introducing and then filling artificial gaps within individual images, and we developed a technique for using the distance image to derive an estimated error associated with each filled pixel.