Global bare ground gain from 2000 to 2012 using Landsat imagery

Qing Ying^{a,*}, Matthew C. Hansen^a, Peter V. Potapov^a, Alexandra Tyukavina^a, Lei Wang^a

Stephen V. Stehman^b, Rebecca Moore^c and Matthew Hancher^c

^a Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA

^b Department of Forest and Natural Resources Management, State University of New York, Syracuse, NY 13210, USA

^c Google, Mountain View, CA, USA

* Corresponding author.

E-mail address: qying@umd.edu (Q. Ying).

Abstract

Bare ground gain, or vegetative cover loss, is an important component of global land cover change resulting from economic drivers such as urbanization and resource extraction. In this study, we characterized global bare ground gain from Landsat time series. The maps were then used to stratify the globe in creating a sample-based estimate of global bare ground gain extent, land cover / land use outcomes, and associated uncertainties from 2000 to 2012. An estimated total of 93,896 km² ($\pm 9,317$ km² for 95% confidence interval) of bare ground gain occurred over the study period. Human-induced bare ground gain accounted for 95% of the total and consisted of the following components: 39% commercial and residential development, 23% resource extraction, 21% infrastructure development, 11% transitional, and 1% greenhouses. East Asia and the Pacific accounted for nearly half of all global bare ground gain area (45%), with China alone accounting for 35% of global gain. The United States was second to China, accounting for 17% of total bare ground gain. Land cover / land use outcomes of bare ground gain varied between regions and

countries, reflecting different stages of development and the possible use of bare ground gain as an indicator of economic activity.

Keywords: Bare ground gain; Urbanization; Resource extraction; Area estimation; Uncertainty quantification; Accuracy assessment; Stratified sampling; Land cover and land use change; Landsat

1. Introduction

Land cover dynamics have been recognized as a key component of global environment change (Foley, 2005) and an important driver to a wide range of ecological, hydrological and climatic processes (Vitousek et al., 1997). Examples include deforestation, urbanization, and agricultural expansion, among others. Each land dynamic has a unique set of socio-economic-political drivers and resulting impacts on the earth system. The complete removal of vegetation due to land use changes such as the expansion of human settlements, open pit mining and infrastructure development represents an extreme land cover transition. This dynamic, which we term bare ground gain (Hansen et al., 2014), includes all vegetation cover loss. Though the conversion of vegetation cover to bare ground cover accounts for a small proportion of global land area, it merits attention for the following reasons. First, bare ground cover is a fast growing land cover type associated with increasing population growth and urbanization. More than half of the world's 7 billion people live in urban areas with an additional 1.3 million urbanites per week (IPCC, 2015). The expected increase in urban area in the first three decades of the 21st century is projected to be greater than the cumulative urban expansion of all human history; urban area is growing on average twice as fast as urban population (IPCC, 2015). Second, bare ground gain completely alters the structure and functioning of ecosystems (Alberti, 2005; Foley et al., 2005; Vitousek et al., 1997)

due to the permanent or semi-permanent removal of vegetation, resulting in lower land carbon storage (Seto et al., 2012), reduced landscape evapotranspiration (Moran et al., 1996; Shukla et al., 1990), and increased surface albedo (Bonan et al., 1992); these effects impact ecosystem functions such as biogeochemical cycling of carbon and water, energy exchange, and biodiversity (Grimm et al., 2008; Kalnay and Cai, 2003). Lastly, bare ground gain merits attention because bare ground cover exhibits complicated spatio-temporal dynamics related to variations in driving forces (Ellis and Ramankutty, 2008; Lambin et al., 2003). For example, economic, demographic and institutional factors and their interactions not only drive local expansion of residential clusters, transportation infrastructure and industrial development (Seto et al., 2012), but also influence land-use change elsewhere (DeFries et al., 2010; Geist and Lambin, 2002). An improved understanding of bare ground gain and associated drivers at a global scale can inform future climate change mitigation actions and human adaptation strategies (IPCC, 2015).

Remotely sensed airborne/satellite data have long been used for bare ground-related land cover themes (Friedl et al., 2002; Gong et al., 2013; Hansen et al., 2003, 2011; Homer et al., 2004; Loveland et al., 2000) and bare ground gain (Hansen et al., 2014) monitoring at various spatial scales. However, definitions related to bare ground cover vary, e.g., continuous versus discrete (Hansen et al., 2003, 2011; Zhan et al., 2002), permeable versus impervious surface (Gong et al., 2013; Homer et al., 2004; Xian and Homer, 2010), natural exposed soil, rock or sand versus anthropogenic built-up environments (Friedl et al., 2002; Loveland et al., 2000). Hansen et al. (2014) featured bare ground gain in strictly cover terms (i.e., the change from vegetation to non-vegetated state) whereas other studies (Schneider, 2012; Sexton et al., 2013; Wang et al., 2010, 2012; Xian and Homer, 2010) characterized impervious surface (non-evaporating, non-transpiring imperviousness) or contiguous patches of human built-up area as an indicator of urban land use

extent. Stable night light data from DMSP-OLS (the U.S. Air Force Defense Meteorological Satellite Program-Operational Linescan System sensors) and VIIRS (Visible Infrared Imaging Radiometer Suite) have been used to map the extent and growth of lighted areas associated with urban areas with limitations related to per capita energy use and satellite intercalibration (Small et al., 2005; Zhang and Seto, 2011). Land uses such as open pit mining or quarries fall outside of most natural bare land, impermeable surface or built-up land cover classification legends. In our definition, bare ground cover includes natural and anthropogenic non-vegetated land surfaces. Hence, we characterize bare ground gain as complete land conversion from vegetative cover to non-vegetative cover.

Following previous research by Hansen et al. (2003,2011,2014), bare ground gain is defined as a process of land cover change featuring complete or semi-permanent (at least 3 years) clearing of vegetative cover (i.e., a pixel experiences an increase in bare ground cover of over 50% or experiences a transition to 80% or greater bare ground cover) by either human or natural-induced disturbances. This land cover dynamic has the advantage of being defined without regard to land use or attributes such as imperviousness, allowing for a more efficient algorithmic implementation. As a result, a generic spectral signature of vegetated to non-vegetated state can be characterized and extended over large areas; herein, we employ 30m spatial resolution Landsat data to map bare ground cover and bare ground gain at the global scale.

Human-induced bare ground gain is due to residential, commercial, industrial, and transportation development as well as excavation and infrastructure related to resource extraction. Naturally-induced bare ground gain results in the exposure of rubble, lava, sand bars and other features caused by, for example, landslides, volcanic eruptions, and river meanders. All human and naturally-induced bare ground gain land changes are distinct from agricultural or forestry land

use practices that include ephemeral bare ground cover states. While an agricultural fallow or intensive tree harvest may in some cases result in an extended period of bare ground exposure, these cases are not included in the permanent or semi-permanent transformation of vegetated land covers to bare ground dominated land covers. We identified six bare ground gain land cover / land use outcomes, five of which are human-induced: resource extraction, infrastructure development, commercial/residential built-up, transitional bare gain and greenhouses. Natural bare ground gain dynamics were considered as a single dynamic. See Methodology section 3.2 for the formal list of land cover / land use outcomes.

Recent developments in optical remote sensing hold tremendous promise for systematic monitoring of global bare ground cover and gain. Hansen et al. (2003) employed a regression tree model to produce the first global continuous fields of percent bare ground cover from Moderate Resolution Imaging Spectroradiometer (MODIS) data. The open access to the USGS Landsat data archive (Woodcock et al., 2008) enabled further improved thematic representation of bare ground cover and gain at Landsat-scale over the conterminous United States (CONUS) through Web-Enabled Landsat Data (WELD) (Hansen et al., 2014). Challenges to using Landsat data over large areas include unequal observation coverage due to cloud cover, scan-line corrector off (SLC-off) gaps, and variation in the number of Landsat acquisitions (Kovalskyy and Roy, 2013). Results for CONUS (Hansen et al., 2014) yielded a binary description (gain/no gain) of bare ground gain with user's and producer's accuracies of 62% and 75%, respectively, when adjacent errors were excluded. Common errors in bare ground gain for land use change mapping include commission errors over croplands with extended fallows and omission errors for developed areas within semi-arid landscapes (Hansen et al., 2014).

The study presented here characterizes bare ground gain using Landsat time-series inputs and employs the resulting maps to stratify the global land surface into areas of likely bare ground gain. A sample of over 5000 pixels was selected to assess map accuracy, estimate area of bare ground gain, and estimate the proportion of area gain attributable to different bare ground gain dynamics. Probability-based sampling methods are regularly used in forest inventories and in remote sensing applications for assessing map accuracy and estimating area (Olofsson et al., 2014). Due to inherent errors in land cover change maps derived from remotely sensed images, change areas obtained from pixel counting are likely biased (Olofsson et al., 2013). Instead, unbiased estimates of area of land cover change area may be produced using the sample and reference classification of change (Stehman, 2013). The map provides a way to target the sampling to a class of primary interest, in this case bare ground gain, via strata that provide sampling efficiencies greater than achieved by simple random or systematic approaches (Broich et al., 2009). For example, cities, towns and settlements account for less than 1% of global land area (Schneider et al., 2010). Given the relative rarity of such land uses compared to the overall land surface, accurate area estimation is a challenge. Here, we are interested not only in mapping the class, but its increase over time, which is an even rarer land theme. To do so, we employ a change map of sufficient quality to construct strata that allow intensifying the sample into potential change areas. The mapped change strata serve to target the theme of interest and greatly reduce the standard error in providing unbiased estimates of change.

Another advantage of a sample based approach is the ability to determine additional contextual information such as land cover state prior to change, land use drivers and the timing of changes for the sampled pixels (Tyukavina et al., 2015). In the study of Tyukavina et al. (2015), forest cover loss was assessed, with all sample pixels interpreted for forest type (natural or managed),

allowing for development of a more complete narrative of forest cover loss. We build on this approach in quantifying global bare ground gain dynamics. Specifically, we estimate the area of land converted to a non-vegetated state and attribute this dynamic to a set of bare ground gain land cover / land use outcomes.

2. Data

We employed Landsat mosaics from the research of global forest dynamics of Hansen et al. (2013) as inputs for mapping. In their study, 654,178 growing season Landsat 7 Enhanced Thematic Mapper Plus (ETM+) scenes from a total of 1.3 million stored in Google Earth Engine cloud platform were analyzed to form a global seamless composite dataset available annually from 1999 to 2012. Annual seamless composite images consist of Landsat 7 ETM+ per band median reflectance values of all cloud/shadow free growing season observations. Growing seasons were defined using MODIS phenology data and all Landsat observations processed therein. Cloud, shadow and water were screened for every pixel using a series of quality assessment models. Viable land observations were normalized to top of canopy reflectance for spectral ETM+ Red (0.631-0.692 μm), Near-Infrared (NIR 0.772-0.898 μm) and two Shortwave Infrared bands (SWIR 1.547-1.749 μm and 2.064-2.345 μm) (Potapov et al., 2012, 2015).

Two additional normalized difference band ratios were derived for their ability to facilitate the characterization of land conversion from vegetation cover to bare ground cover. Annual growing season minimum greenness (Normalized Difference Vegetation Index) values were computed as was a normalized difference built-up index (NDBI) $(\text{SWIR } 1.6 \mu\text{m} - \text{NIR}) / (\text{SWIR } 1.6 \mu\text{m} + \text{NIR})$ (Zha et al., 2003) using the median growing season data.

3. Methodology

3.1 Global bare ground gain characterization

A host of land cover change methods exist (Coppin et al., 2004; Singh, 1989). In this study, we characterize bare ground gain both directly as a class (see section 3.1.1) and indirectly using post-characterization comparison of annual percent bare ground cover layers (see section 3.1.2). The work flow of global bare ground gain characterization using the direct and indirect methods is illustrated in Fig.1. Annual Landsat mosaics were inputs for both direct bare ground gain mapping (Fig.1 left column) and indirect bare ground mapping (Fig.1 right column). Classification trees were used for directly mapping bare ground gain from 2000 to 2012. For the indirect mapping of bare ground gain from 2000 to 2012, we used regression trees to map annually bare ground cover from 1999 through 2012 and then developed rules for determining from these annual maps whether bare ground gain had occurred.

Training data for directly mapping bare ground gain were delineated as gain/no gain labels by interpreting stacks of Landsat images supplemented by high resolution images from Google EarthTM (section 3.1.1.2). Additional ancillary data to add to the bare ground gain training in the USA included imperviousness change data from the National Land Cover Dataset (NLCD) (Jin et al., 2013; Xian et al., 2011) and bare ground gain data from the WELD dataset (Hansen et al., 2014). Training data were related to a set of multi-temporal metrics (see metric explanation in section 3.1.1.1), or statistical derivatives of the annual Landsat composite images, using a classification tree. Classification tree models were applied in an active learning mode, with global bare ground gain products iteratively created until a final, acceptable result was achieved based on expert interpretation (section 3.1.1.3). The final result was a classification of pixels having experienced bare ground gain from 2000 to 2012.

180 Annual percent bare ground cover layers were created using a regression tree model, equivalent
181 to the MODIS Vegetation Continuous Field (VCF) and Landsat-based WELD products (Hansen
182 et al., 2003, 2011). For these antecedent studies, time-series data sets were related to fractional
183 cover training data using regression trees in order to estimate per pixel percent bare ground. Both
184 products were used as training data in this study per the approach of Hansen et al. (2014). Post-
185 characterization rules of bare ground increase $\geq 50\%$ over a consecutive three year period were
186 employed to classify bare ground gain from 2000 to 2012 using the series of annual bare ground
187 cover maps (section 3.1.2).

Fig.1. Work flow of global bare ground gain characterization. Annual Landsat mosaics are inputs. Two bare ground gain characterization methods are shown: direct bare ground mapping and indirect bare ground gain mapping. Colored boxes identify training data, algorithm and output maps applied to each mapping method. Orange boxes at bottom show the structure of mapped strata for validation and sample-based area estimation. Note that mapping is not used for area estimation, but as strata for sample-based area estimation.

3.1.1 Direct bare ground gain mapping

3.1.1.1 Multi-temporal metrics extraction

The spectral profiles of bare ground gain pixels vary greatly because of the complicated spatio-temporal characteristics of bare ground extent and change. For example, the spectral profiles of conversions differ by the timing, duration, magnitude, and reference start and end states (Fig.2). From the annual composites and indices, we computed multi-temporal metrics meant to capture interannual spectral variations suitable for characterizing bare ground gain. The method of deriving metrics from satellite time series has been used in land cover change studies of forest loss and bare ground gain (Hansen et al., 2008). Five groups of per band metrics were created: (1) earliest and latest cloud-free observation composites and median value of 3 earliest/latest observations; these metrics represent interval endpoints and their respective differences (earliest – latest; median 3 earliest – median 3 latest) are a generic indicator of change; (2) statistical measures of minimum/maximum annual values and the median value of 3 highest/lowest observations from the multi-year profiles; these metrics represent extreme values within the annual inputs; their differences (max - min, median 3 max - median 3 min) are indicators of change magnitudes; (3) average of all years of annual inputs; (4) slopes of linear regression models of annual input values versus year; (5) maximum change for both increase and decrease segments and corresponding endpoints in respective annual profiles. A total of 21 metrics were created per band and ratio and annual bare ground cover inputs, resulting in a set of 147 feature metrics (Table 1).

Table 1. Summary of multi-temporal metrics extracted from Landsat 7 temporal profiles

Multi-temporal metrics extraction	Names
7 Features	Four spectral bands: median growing season Red, NIR, SWIR 1.6 μm , SWIR 2.2 μm Two indices bands: minimum growing season NDVI, median growing season NDBI

	Percent bare ground cover
	Six temporal metrics: earliest, latest, median 3 earliest, median 3 latest, earliest - latest, median 3 earliest - median 3 latest
	Six time-integrated metrics: maximum, minimum, median of 3 maximum, median of 3 minimum, maximum - minimum, median of 3 maximum - median of 3 minimum
21 Multi-temporal metrics per band	Mean of all years
	Slopes of annual input values versus year
	Six maximum bare ground gain increase / decrease metrics: start point, end point and magnitude of maximum increase and decrease from annual profiles
Total	147 feature metrics

Fig.2. Band profiles of bare ground gain pixels from (a) a coal mining site in Alabama, USA, centered at 33°53'20"N, 87°32'36"W and (b) a power plant in North Carolina, USA, centered at 34°50'32"N, 79°44'8"W, above which are Landsat false color composites of SWIR (1.6 μm), NIR and RED bands corresponding to the start, change and end year.

3.1.1.2 Global training dataset of bare ground gain

The most definitive way to obtain reference information for training data to classify bare ground gain would be to collect field data over time, an impractical option for global studies. Google Earth data offer high-resolution historical images that cover roughly twenty percent of the Earth's landmass and more than a third of the human population (Potere, 2008) with coverage improving over time. A wide variety of plug-ins and tools built upon Google Earth facilitate global land change research (Fritz et al., 2009). Because of the efficiency in visualizing the trajectories of historical change of areas of interest, Google Earth has been exploited in a wide range of land cover change studies (Hansen et al., 2011). In this study, Google Earth was employed in both training and validation, with Landsat annual composites used for both activities.

A global training data set for bare ground gain classification was developed through visual interpretation and on-screen delineation of gain and no gain classes. A three feature false color composite consisting of the median value of 3 earliest, latest and maximum observations in percent bare ground cover was used in training data delineation. Start and end date images of Landsat false color composites and Google Earth time slider imagery were also used to facilitate interpretation.

A particular focus of training data derivation is precise labeling of boundary pixels which is important for improving the accurate characterization of mixed pixels (Hansen, 2012; Hansen et al., 2014). In active learning mode, training data were delineated and iterated until little or no improvement in successive versions was observed. Metropolitan areas such as Washington D.C., Dallas, Manchester, Moscow, Istanbul, Lagos, Dubai, Shanghai, Beijing, Manila and Melbourne were trained in order to represent a wide range of urbanization dynamics. Oil and gas drilling and mining sites were sampled across the globe, from China to Russia and the USA and Canada to Peru and Australia. In addition to adding bare ground gain training data, commission errors were

targeted, largely in fallowed agriculture lands. Training data were regionally complemented by adding gain and no gain pixels outside our delineated polygons where WELD bare ground gain layer and NLCD percent developed imperviousness change layer agreed (>10% imperviousness increase regarded as NLCD gain, others as NLCD no gain) for the purpose of enlarging no gain training. Exclusion of ephemeral changes, even sometimes lasting more than one year, such as post-clearing forest patches, agricultural fallows, and seasonal changes of shallow water, was sought in the delineation of training data. Pixels with noise patterns associated with inter-annual bare ground cover variations, such as high latitude snow/ice cover, were also added to the no bare ground gain class. The effort to include pixels of likely false positive bare ground gain in the training data is critical to improving the user's accuracy of the final product. The global training data set for bare ground gain classification totaled 27,471,758 Landsat pixels, of which 188,860 were gain pixels and 27,282,898 no gain pixels.

3.1.1.3 Classification tree model for bare ground gain detection

Decision trees have been widely used in classifications of large area land cover extent and change (Hansen et al., 2000, 2014; Potapov et al., 2015; Schneider et al., 2010; Weber et al., 2009). Tree models have been shown to perform better in urban area land cover change detection compared to support vector machines (SVM) (Schneider, 2012). Decision trees are a supervised classification algorithm that recursively partitions a training data set into more homogeneous subsets with regard to class membership until certain conditions are met (Breiman et al., 1984). A bagging technique was employed to reduce overfitting in individual tree models (Breiman, 1996). Considering the number of tree models and the structure of training data set, we used equal sampling rate (20%/20%) of delineated training for gain/no gain classes and unequal rate (10%/0.1%) for complementary training data obtained from existing products. Seven bagged trees

(Fig.1 left blue box) were built via repeatedly sampling the training data (Potapov et al., 2015) and then applied to predict per pixel bare ground gain probability (Fig.1 left purple box). A pixel was labeled as bare ground gain when the per pixel median value of all tree models was greater than 50% (Fig.1 left red box).

3.1.2 Indirect bare ground gain mapping

The indirectly mapped bare ground gain was produced based on the concept of post-characterization mapping using the consecutive annual bare ground cover layers as inputs. A pixel was defined as bare ground gain if it had experienced an increase in bare ground cover of at least 50% within a consecutive three year period. Specifically, as shown in equation (1), we first extracted a time sequence of bare ground cover ($\hat{x}_t$) that was stable within a consecutive three year period for a pixel from its annual percent bare ground cover time series (x_t). Then we mapped the pixel as bare ground gain if the maximum increase of the time sequence $\hat{x}_t$ was equal to or larger than 50% as shown in equation (2). To meet the consecutive three year constraint, we used percent bare ground cover data starting from 1999 for calculating stable bare ground cover in 2000. We only used year t-1 and t percent bare ground cover to define stable bare ground cover in 2012 as we did not have year 2013 data at the time of the study.

$$\hat{x}_t = \text{median}(x_{t-1}, x_t, x_{t+1}), \text{ if } \max(x_{t-1}, x_t, x_{t+1}) - \min(x_{t-1}, x_t, x_{t+1}) \leq 30\% \quad (1)$$

$$\text{Pixel } i \text{ is classified as bare ground gain, if } \Delta\hat{x} \geq 50\% \quad (2)$$

3.2 Sampling design for accuracy assessment and area estimation

We implemented a stratified random sampling design for the purpose of providing area estimates of bare ground gain and for assessing the accuracy of the bare ground gain map. Land cover change typically represents a small fraction of the land surface at decadal time scales,

meaning that any such assessment requires special attention to the change class. Khorram (1999) pointed out the difficulty in assessing land cover change map accuracy due to the difficulty in quantifying false negatives. Such errors of land change omission are found within a relatively huge stratum of mapped non-change. Considering that built-up land as a land category consists of only 0.5% of the land surface (Schneider et al., 2009), accurately mapping its change within a global context is a challenge. The mapped bare ground layers enable the creation of a set of strata designed to target likely areas of bare ground gain and associated errors.

The strata used in the sampling design were defined by a cross-classification of five climate domains with five map-based feature classes. The five climate domains were tropical, subtropical, temperate, boreal and polar according to the eco-climate zones of the Food and Agriculture Organization (Simons et al., 2001; downloaded from [FAO Map Data Global Ecological Zones](#)). The five feature classes within each domain were defined by different combinations of the classifications produced by the direct and indirect bare ground gain products (Fig.1 bottom orange boxes). The first map feature consisted of pixels mapped as gain in both the directly and indirectly mapped products (“jointly mapped” bare ground gain as shown in red in Fig.3 strata inset). This layer represented the highest confidence feature class for bare ground gain. The “directly mapped” feature class was created using pixels identified as bare ground gain by only the direct mapping algorithm (as shown in yellow in Fig.3 strata inset), and the “indirectly mapped” feature class was created using pixels identified as bare ground gain only by the post-characterization comparison (as shown in blue in Fig.3 strata inset). A fourth “buffer” feature class was defined by all pixels within 90m of any pixel included in any of the first three map feature classes (as shown in grey in Fig.3 strata inset). This buffer class was designed to increase the likelihood of sampling bare ground gain omission errors, which presumably would be more likely to occur in areas proximate

to mapped bare ground gain. A fifth “mapped no gain” feature class consisted of all the terrestrial pixels not identified by the aforementioned four feature classes (as shown in white in Fig.3 strata inset). The five climate domains combined with the five map features resulted in a total of 25 strata, and the area covered by these 25 strata is the population to which inferences from the sample data apply (Table 2). Fig.3 shows an example of the strata structure.

Fig.3. Example of stratification for Brawley, California (a 10 km by 6.5 km extent centered at $32^{\circ}59'5''N$, $115^{\circ}30'33''W$), within the subtropical domain. Strata are shown on the right image where red is jointly mapped gain (Stratum 6), yellow is directly mapped (Stratum 7), blue is indirectly mapped (Stratum 8), and dark gray is the 90m buffer (Stratum 9); white is mapped no gain (Stratum 10).

Globally, 5750 sample pixels (Fig. 4) were selected using the stratified random design (Table 2) following “good practice” guidance on the use of probability-based sampling for area estimation (Olofsson et al., 2014). A sample is represented by a Landsat 30-m spatial resolution pixel. We first selected 4250 sample pixels with 4000 in bare ground gain strata and buffer strata and 250 in no gain strata. The allocation of the 4250 sample size per climate domain was approximately proportional to bare ground gain area of each domain: tropical, subtropical, and temperate domains were each initially allocated 1050 sample pixels and the boreal and polar domains were each allocated 550 pixels. Within each climate domain, we selected equal number of sample pixels from

each of the four gain strata and 50 sample pixels from the no gain stratum. Initial results from the 4250 sample points revealed that the majority of bare ground gain estimation uncertainty was contributed from the buffer strata within the tropical, subtropical and temperate climate domains. Consequently, to reduce the standard error of the area estimate we added 500 sample pixels to each buffer stratum of the tropical, subtropical and temperate climate domains.

Reference information was derived via visual interpretation of thumbnail annual false color composites of SWIR (1.6 μm), NIR and RED for 3km by 3km (101 x 101 Landsat pixels) block centered on the sample pixel location. The thumbnails provide landscape context to facilitate interpretation of the sampled pixel. Sampled pixel boundaries were also projected onto Google Earth to facilitate interpretation for locations with very high spatial resolution imagery. When interpretation ambiguity due to poor geolocational agreement between Landsat pixels and Google Earth imagery occurred, priority was given to the interpretation of the Landsat time-series. The primary task was to assign if bare ground gain occurred or not. For reference sample pixels labeled as bare ground gain, the change year, antecedent land cover/use and attributes of bare ground gain drivers were interpreted and recorded.

Based on the USGS (United States Geological Survey) Anderson land classification scheme (Anderson, 1976), we define six classes of bare ground gain land cover and land use outcomes that constitute the dominant drivers of bare ground gain:

- (1) Resource extraction: the removal of vegetation to gain access to mineral or energy commodities, including exploration and extraction activities for iron ore, limestone, gold, copper, zinc and other minerals, and coal, petroleum, natural gas and other fossil fuels;

- (2) Infrastructure development: the construction of transportation and energy infrastructure including highways, railways, airports, container terminals, ferries, dams, power stations, pipelines, and manufacturing plants;
- (3) Commercial/residential built-up: structures related to human occupation, such as residential homes and office buildings and including associated features like driveways and parking lots;
- (4) Transitional bare gain: areas which are in transition from one land use activity to another. For example, the conversion of agricultural land to residential land use may exist in a bare ground cover state for an extended period and in fact not reach completion, for example the cancellation of a construction project due to changes in market conditions or financing. Spoil dumps and landfills which are also transitional in nature are also considered in this category;
- (5) Greenhouses: a special category characterized by impervious roofing materials made of glass, polycarbonate or plastic panels used in the perennial production of fresh vegetables and fruits. This category is selected because greenhouses have similar or same spectral reflectance as other bare ground expansion areas;
- (6) Natural bare ground gain: all naturally caused long term exposure of bare soil, sand and rock (for example by river meanders), dry salt flats, earthquake craters, and volcanic debris.

Table 2. Allocation of sample size to strata formed by the cross-classification of climate domains with map feature classes.

					Stratum
Stratum ID	Climate domains	Sample size	Map features	Sample size	area (km ²)
1	Tropical	1550	Jointly mapped	250	1,897

2			Directly mapped	250	16,000
3			Indirectly mapped	250	15,425
4			Buffer	750	304,769
5			Mapped no gain	50	57,240,695
6			Jointly mapped	250	3,703
7			Directly mapped	250	17,471
8	Subtropical	1550	Indirectly mapped	250	27,975
9			Buffer	750	412,966
10			Mapped no gain	50	22,284,674
11			Jointly mapped	250	5,011
12			Directly mapped	250	19,359
13	Temperate	1550	Indirectly mapped	250	28,319
14			Buffer	750	455,944
15			Mapped no gain	50	27,624,553
16			Jointly mapped	125	664
17			Directly mapped	125	3,279
18	Boreal	550	Indirectly mapped	125	1,110
19			Buffer	125	60,380
20			Mapped no gain	50	18,923,617
21			Jointly mapped	125	26
22			Directly mapped	125	360
23	Polar	550	Indirectly mapped	125	771
24			Buffer	125	24,765

Fig.4. Stratified random sample with 25 strata defined from the cross classification of five climate domains with five map feature classes.

3.3 Accuracy assessment, area estimation and uncertainty quantification

We employ the method of Olofsson et al. (2013) that utilizes the error matrix obtained from validation to produce a stratified estimator of bare ground gain area based on the reference classification of bare ground gain. Bare ground gain area was estimated for each of the 25 strata within a region of interest, and the stratum totals were then summed to obtain the total estimate for a designated region such as global, continental or country areas. The bare ground gain area in stratum h is estimated as

$$\hat{G}_h = A_h p_h \quad (3)$$

where A_h is the total area in stratum h , $p_h = n'_h/n_h$ is the sample proportion of pixels interpreted as bare ground gain in stratum h , n'_h is the number of sample pixels interpreted as bare ground gain, and n_h is the number of sample pixels allocated to stratum h . Then the total bare ground gain area is estimated by summing the area estimates over all 25 strata:

$$\hat{G} = \sum_{h=1}^{25} \hat{G}_h \quad (4)$$

We quantify the uncertainty of the area estimates in the context of design-based inference framework in which the uncertainty associated with the estimator is defined as the variability of the estimates over the set of all possible samples that could have been obtained for the chosen sampling design and population sampled. The standard error for bare ground gain area estimate is

$$S(\hat{G}) = \sqrt{\sum_{h=1}^{25} A_h^2 \frac{p_h(1-p_h)}{n_h-1}} \quad (5)$$

A 95% confidence interval for bare ground gain area is

$$\hat{G} \pm 1.96 \times S(\hat{G}) \quad (6)$$

For map accuracy assessment, we use the estimation equations of Stehman (2014) because the strata are different from the map classes. Detailed descriptions of the formulas for user's, producer's and overall accuracy are presented in the Appendix.

Although we employed climate domains as strata in the sampling design, we are also able to report area estimates for other geographic regions if there is sufficient sample size within the region to assure a relatively low standard error. For example, we report results for continents/sub-continents and countries where sampling densities were sufficient to produce a reasonable standard error (e.g. China, the United States of America, Brazil, Canada, Russia, India, Mexico, Spain, Turkey, Australia). The hierarchy of continents/sub-continents is based on the geographic

groupings of World Bank (Fig.5) that include all countries and economies with populations of more than 30,000 (World Bank, 2016a). The World Bank has defined geographic regions for a wide range of development indicators and we use them to disaggregate area estimates of bare ground gain and associated land cover / land use outcomes.

Fig.5. Boundaries of geographic regions from the World Bank for regional bare ground gain area reporting. The sample size located in each region is: East Asia and Pacific 1444, Europe and Central Asia 1416, North America 987, Latin America and the Caribbean 608, South Asia 486 and Sub-Saharan Africa 411, Middle East and North Africa 377. There are 21 sample pixels falling outside of the boundaries that we excluded from area estimation.

4. Results

4.1 Global bare ground gain characterization and stratification results

The area comprising the population from which the sample was selected is global terrestrial area between 80° N to 60° S. The mapped bare ground gain and associated buffer area covered

1,400,193 km² or approximately 1% of global land area. The mapped global bare ground gain area for 2000-2012 from the direct classification map totaled 67,771 km² (0.05% of global land area). Two example bare ground gain landscapes of resource extraction and urbanization from the direct classification map are illustrated in Fig.6. Indirect mapped area totaled 85,172 km² (0.06% of global land area). To visualize the global distribution of bare ground gain, we assigned the sample-based stratum mean bare ground gain to each pixel in each stratum and then aggregated the result to a 0.05 degree spatial resolution (Fig.7a). Bare ground gain patterns are largely related to urban expansion, for example the metropolitan agglomerations of Houston-Dallas-San Antonio in Texas (Fig.7b) and the Beijing-Tianjin-Tangshan corridor in the North China Plain (Fig.7c). Infrastructure development is also readily visible in the bare ground gain maps. For example, expanding human settlements are connected by the growing transportation networks in Northern China (Fig.7c). Likewise, the construction of Dulles International airport in northern Virginia has driven concentrated development of satellite cities in the Washington, D.C. metro area (Fig.7b) (Song et al., 2016). Resource extraction is present in Fig. 7b, including the East Texas oil field, the second-largest oil field in the U.S. outside of Alaska, and mountaintop removal mining in the Appalachian Plateau of West Virginia and Kentucky. These land cover / land use outcomes of bare ground gain were attributed to sample points during the validation procedure (Fig.8) to quantify the proportions of human-induced and natural bare ground gain.

The direct and indirect bare ground gain maps corresponded well with visually interpreted bare ground gain. We disaggregated errors by evaluating map accuracy for different combinations of map strata evaluated as the bare ground gain class (Table 3). For pixels identified as bare ground gain in both direct and indirect bare ground gain maps (namely jointly mapped pixels), commission error was just over 10% and omission error was nearly 90%. Adding the directly mapped bare

ground gain to the jointly mapped area increases commission error to 33% and reduced omission error to 50%. Following Hansen et al. (2014), we define adjacent errors as any omission or commission errors that were found within a 1-pixel distance from directly mapped bare ground gain. Adjacent errors are more likely related to geolocation and / or mixed pixel issues than errors that are isolated from correctly mapped change (Hansen et al., 2014). When adjacent errors were excluded, in other words the adjacent errors were not counted as error, the user's accuracy increased to 92% and producer's increased to 66%. In our directly mapped bare ground gain, 54% of omission errors were within one pixel of a labeled gain pixel and an additional 18% of omission errors were within two pixels of a labeled gain pixel. The addition of indirectly mapped bare ground gain further increased commission error to 62% but reduced omission error to 44%. Because we did not find any gain sample pixels from the mapped no gain strata, the inclusion of the buffer as mapped bare ground gain increased commission error to 93% but had no omission error, verifying our original purpose to employ the buffer strata for targeting false negatives. These results show the expected tradeoff between commission and omission errors as the area of mapped bare ground gain increases.

The accuracy estimates of both direct and indirect bare ground gain maps also varied by climate domains (Table 4). As expected, the overall accuracy was high for all climate domains because of the very small proportions of bare ground gain area. The general pattern of accuracy over the climate domains was that user's and producer's accuracies of bare ground gain were lower in the polar and tropical domains with higher accuracies in the subtropical, temperate and boreal domains. Adding directly mapped bare ground gain to the jointly mapped gain gained the largest increase in producer's accuracy of bare ground gain class in boreal domain, but introduced the most reduction in user's accuracy in polar domain.

Fig.6. Landsat composites and directly classified bare ground gain: (a) Petroleum drilling in rural Texas centered at $30^{\circ}24'33''\text{N}$, $101^{\circ}02'46''\text{W}$; (b) Urbanization in Beijing, China centered at $39^{\circ}48'01''\text{N}$, $116^{\circ}33'07''\text{E}$. The first and last composites are Landsat false color composites of SWIR ($1.6\ \mu\text{m}$), NIR and RED bands. Bare ground gain is shown in red on images of right panel with background displayed in the false color composite of median_max, median_latest and median_earliest in percent bare ground cover as RGB.

Fig.7. Percent bare ground gain aggregated from Landsat strata to 0.05° resolution in lat/lon geographic projection.

476 Fig.8. Examples of characterized bare ground gain land cover / land use outcomes in a 930m ×
 477 930m (31 × 31 Landsat pixels) block centered at sample points as shown in red square: (a)
 478 Resource extraction, coal mine in Queensland, Australia centered at 21°29'29"S 148°23'12"E; (b)
 479 Infrastructure development, airport runway in Chongqing, China centered at 29°42'29"N
 480 106°38'35"E; (c) Commercial/residential built-up, shopping plaza in Oregon, U.S. centered at
 481 45°0'46"N 122°59'54"W; (d) Transitional bare ground gain, bare field next to a building in Rio
 482 Grande do Sul, Brazil centered at 29°51'7"S 50°14'53"W; (e) Natural bare ground gain, river
 483 meander in Assam, India centered at 27°46'45"N 95°39'11"E; (f) Greenhouses in Antalya, Turkey
 484 centered at 37°0'16"N 30°50'49"E. Images from left to right column are: 1), 2) reference Google
 485 Earth time one and time two imagery; 3),4) Landsat start and end date composites of SWIR (1.6
 486 μm), NIR and RED bands as RGB; 5) Direct bare ground gain classification results with gain in
 487 yellow and no gain in black.

488

489 Table 3. Accuracy and uncertainty ($\pm$ margin of error for 95% confidence interval) estimates of the bare ground gain and no gain classes
490 for different combinations of map features. Each row represents one of the objective bare ground gain maps and the accuracy estimates
491 represent the global result combining the sample data for all five climate domains.

Map Feature	User's		Producer's		Overall
	Gain	No gain	Gain	No gain	
Jointly mapped	87.7 (± 2.4)	99.9 (± 0.01)	10.8 (± 1.1)	100.0 (± 0.00)	99.9 (± 0.01)
Jointly mapped + Directly mapped	66.9 (± 2.7)	99.9 (± 0.01)	48.5 (± 4.8)	99.9 (± 0.00)	99.9 (± 0.01)
Jointly mapped + Directly mapped + Indirectly mapped	37.5 (± 1.7)	99.9 (± 0.01)	56.3 (± 5.5)	99.9 (± 0.00)	99.9 (± 0.01)
Jointly mapped + Directly mapped + Indirectly mapped + Buffer	6.6 (± 0.6)	100.0 (± 0.00)	100.0 (± 0.0)	99.1 (± 0.01)	99.1 (± 0.01)

492

493 Table 4. Accuracy and uncertainty ($\pm$ margin of error for 95% confidence interval) estimates of the bare ground gain and no gain classes
494 disaggregated by climate domains for different combinations of map features.

Map Feature	Climate domains	User's		Producer's		Overall
		Gain	No gain	Gain	No gain	
Jointly mapped	Tropical	82.4 (± 4.7)	99.9 (± 0.01)	6.9 (± 1.4)	100.0 (± 0.00)	99.9 (± 0.01)
	Subtropical	88.4 (± 4.0)	99.8 (± 0.03)	9.6 (± 1.7)	100.0 (± 0.00)	99.8 (± 0.03)

Jointly mapped + Directly mapped	Temperate	87.6 (± 4.1)	99.9 (± 0.02)	13.4 (± 2.4)	100.0 (± 0.00)	99.9 (± 0.02)
	Boreal	95.2 (± 3.8)	99.9 (± 0.01)	15.9 (± 5.9)	100.0 (± 0.00)	99.9 (± 0.01)
	Polar	84.0 (± 6.5)	99.9 (± 0.01)	2.6 (± 2.5)	100.0 (± 0.00)	99.9 (± 0.01)
	Tropical	57.2 (± 5.6)	99.9 (± 0.01)	45.1 (± 8.6)	99.9 (± 0.00)	99.9 (± 0.01)
	Subtropical	73.3 (± 4.7)	99.9 (± 0.03)	44.9 (± 7.8)	99.9 (± 0.00)	99.8 (± 0.03)
	Temperate	69.1 (± 4.8)	99.9 (± 0.02)	52.2 (± 8.9)	99.9 (± 0.00)	99.9 (± 0.02)
	Boreal	66.3 (± 7.2)	99.9 (± 0.01)	66.8 (± 24.2)	99.9 (± 0.00)	99.9 (± 0.01)
	Polar	28.2 (± 7.1)	99.9 (± 0.01)	13.7 (± 13.2)	100.0 (± 0.00)	99.9 (± 0.01)
	Tropical	34.7 (± 3.4)	99.9 (± 0.01)	51.2 (± 9.6)	99.9 (± 0.00)	99.9 (± 0.01)
	Subtropical	37.1 (± 2.9)	99.9 (± 0.03)	53.1 (± 9.1)	99.8 (± 0.01)	99.7 (± 0.03)
Jointly mapped + Directly mapped + Indirectly mapped	Temperate	38.0 (± 3.0)	99.9 (± 0.02)	61.0 (± 10.2)	99.8 (± 0.01)	99.8 (± 0.02)
	Boreal	56.6 (± 5.9)	100.0 (± 0.01)	73.9 (± 26.7)	99.9 (± 0.00)	99.9 (± 0.01)
	Polar	10.1 (± 2.9)	99.9 (± 0.01)	16.4 (± 15.9)	100.0 (± 0.00)	99.9 (± 0.01)
	Tropical	6.7 (± 1.3)	100.0 (± 0.00)	100.0 (± 0.0)	99.4 (± 0.01)	99.4 (± 0.01)
	Subtropical	7.4 (± 1.3)	100.0 (± 0.00)	100.0 (± 0.0)	98.0 (± 0.03)	98.1 (± 0.03)
Jointly mapped + Directly mapped + Indirectly mapped + Buffer	Temperate	6.4 (± 1.1)	100.0 (± 0.00)	100.0 (± 0.0)	98.3 (± 0.02)	98.3 (± 0.02)
	Boreal	5.7 (± 2.1)	100.0 (± 0.00)	100.0 (± 0.0)	99.6 (± 0.01)	99.6 (± 0.01)
	Polar	2.8 (± 2.6)	100.0 (± 0.00)	100.0 (± 0.0)	99.6 (± 0.01)	99.6 (± 0.01)

4.2 Bare ground gain area estimates and uncertainties

Global bare ground gain area based on the stratified random sample was estimated to be $93,896 \pm 9,317 \text{ km}^2$ (95% confidence interval) during the study period, which is about the area of the country of Portugal or the U.S. state of Indiana. Global bare ground gain area was attributed to six land cover / land use outcomes: 39% commercial-residential built-up, 23% resource extraction, 21% infrastructure development, 11% transitional, 1% greenhouses and 5% natural gain. The subtropical climate domain contained 36% of global bare ground gain area, temperate 35%, tropical 24%, boreal 4% and polar 1%.

The estimated area of global bare ground gain was disaggregated by region and country. Regional estimates of bare ground gain area between 2000 and 2012 and associated uncertainty are shown in Fig.9a. Among the seven geographic regions, East Asia and Pacific had nearly half of global bare ground gain area (45%), whereas the land area in this region only accounts for 19% of global land area. North America accounts for about 20% of global bare ground gain area within 14% of global land area. The proportion of bare ground gain area in Europe and Central Asia, Latin America and the Caribbean, South Asia, Sub-Saharan Africa and Middle East and North Africa is 15%, 11%, 4%, 3% and 2%, respectively. This disproportionate distribution is caused by different socioeconomic development rates among regions. Per capita bare ground gain area by region reveals distinct development patterns. For example, per capita bare ground gain area is highest in North America with 54.8 m^2 and lowest in South Asia with 2.3 m^2 . The top ten countries with the largest estimates of bare ground gain area (Fig.9b) account for 74% of global bare ground gain area. China experienced the largest increase in bare ground for the study period, accounting for 35% of global bare ground gain; the U.S. was second with a share of 17%. China and the U.S. differed most markedly in their respective investment in infrastructure (Fig. 9b). Bare ground gain

519 from infrastructure development totaled $11,063 \pm 4,073 \text{ km}^2$ and $2,268 \pm 1,425 \text{ km}^2$, for China and
520 the U.S., respectively (the $\pm$ values represent the 95% confidence interval).

521 Proportional compositions of attributed bare ground gain drivers also vary between regions and
522 countries (Fig.10). The East Asia and Pacific region has the largest total area of bare ground gain
523 from resource extraction due to extensive mining in China and Australia, whereas Sub-Saharan
524 Africa has the highest proportion of bare ground gain due to resource extraction.
525 Commercial/residential built-up and infrastructure development in East Asia and Pacific are
526 largely related to China and its rapid urbanization rate during the study period. North America and
527 the Middle East and North Africa also exhibit high shares of commercial/residential built-up.
528 Much of the Middle East's growth is found in the oil rich Gulf States. South Asia has the greatest
529 share of natural bare ground gain, mainly due to river meanders within Himalayan watersheds.

Fig. 9. Bare ground gain area estimates for 2000-2012 (95% confidence intervals represented by error bars) and composition of land cover / land use outcomes of bare ground gain: a) regional estimates; b) estimates of top ten countries with the largest directly mapped bare ground gain area.

Fig.10. Proportions of estimated bare ground gain area to land cover / land use outcomes: a) by region and b) by country.

5. Discussion and conclusion

Our research delivers the first estimates and uncertainties of global bare ground gain area using Landsat imagery. Bare ground gain represents a very small fraction of the overall global land cover and land use extent and change, making it a difficult theme to quantify. To illustrate this idea and the resulting challenge to area estimation, we can reference tree cover loss for comparison. Tree cover of >25% canopy and 5m tall trees covers nearly 30% of the global land surface (Hansen et al., 2013). Tree cover change as mapped by Landsat occurs within 0.13% of global land area annually (Hansen et al., 2013). It is obvious that in order to establish precise rates of tree cover change, an accurate stratification is required to target this comparatively rare annual dynamic. For bare ground, the challenge is even greater. For example, urban areas account for less than 1% of global land area (Schneider et al., 2010) and their growth over time is vanishingly small compared to tree cover. To accurately estimate the extent of this land dynamic, a mapping capability is required to successfully target this rare land dynamic, something enabled by Landsat data.

Compared to a previous study of Landsat-derived bare ground gain for the continental U.S. (Hansen et al., 2014), mapped bare ground gain of our global study had a relatively higher user's accuracy (67% compared to 42% for the previous continental study) and comparable producer's accuracy (48% compared to 49% for the previous study) even though our study was implemented at a larger scale (global vs country level). For both of these studies, the maps proved inadequate when a "pixel counting" approach to calculating area of bare ground gain was used. However, the use of the map as an indicator of landscapes experiencing bare ground gain facilitated unbiased sample-based area estimates of bare ground gain. We calculated the gain in precision of the bare ground gain area estimates attributable to the stratified sampling design (Cochran, 1977, section 5A.11) as quantified by the ratio of the standard error expected under simple random sampling to

the standard error of stratified random sampling design implemented in our study (both designs have the same sample size). The larger the ratio of standard errors the greater the benefit of the stratification by mapping bare ground gains. The calculation of the standard error ratios in our study is based on the estimated proportion of bare ground gain in each of the 25 strata (combinations of climate domain and map feature class) as these proportions determine the standard deviation for each stratum. Because we found no bare ground gain in the “mapped no gain” strata, we assumed that these strata had 0.01% of bare ground gain because otherwise these five “no gain” strata would have contributed no variance to the overall area estimate. The standard error ratios were 11.5 globally, 14.0 for the tropical domain, 8.7 for subtropical, 9.8 for temperate, 7.7 for boreal, and 12.7 for polar domain. In other words, over eleven times the number of sample pixels selected using the Landsat stratification would have been needed to produce a global bare ground gain estimate with a similar uncertainty using simple random sampling. The sampling approach also enabled the assignment of contextual attributes (e.g. land cover / land use outcomes of bare ground gain) per sample pixel, enriching the overall land cover and land use change narrative. Tyukavina et al. (2015) performed a similar analysis in assigning natural versus managed forest cover loss to samples of pan-tropical forest disturbance. Sample interpretation enables obtaining considerable thematic value-added information not captured by the map.

The preferred outcome and long term objective is to produce land cover change maps that are highly accurate. In this study, the directly classified map outperformed the post-characterization map in terms of accuracy of the bare ground gain class (67% vs 20% for user’s and 48% vs 18% for producer’s), suggesting that our bare ground gain classification approach has the potential to produce more accurate characterization of land change. Mapping bare ground gain indirectly using the percent bare ground cover time series proved less reliable. Serra et al. (2003) and Olofsson et

al. (2013) have demonstrated that post-characterization approaches have limitations in characterizing land cover change. To improve the mapping performance of both approaches, a more sophisticated feature space is needed as well as additional training data focused on regions that performed poorly in terms of map accuracy in this study.

The approach of estimating area from a sample based on visually interpreted bare ground gain offers a cost-effective alternative free of the bias attributable to classification map error. The primary limitation of the sampling approach is that it is far less effective at providing a spatially explicit representation of bare ground gain. While each pixel can be assigned the mean gain of its associated stratum, this would not likely result in a map that was effective for reporting sub-regional area estimates. Consequently, larger sample sizes and possibly more regional/national strata would be required to produce more spatially disaggregated area estimates with adequate precision. The effort to manually interpret a very large sample is the main limitation of sample-based land change estimation when a large area with many sub-regions is of interest.

A larger sample would also better inform possible limitations to the overall method. For example, it could be expected that bare ground gain in semi-arid and arid environments with considerable antecedent fractional bare ground cover might exhibit higher omission error rates. While bare ground gain in arid regions such as urbanization in Dubai, UAE, and semi-arid regions such as mining in Wyoming, USA, were mapped, there may be other dynamics omitted. However, we did not find this to be the case as we did not have any commission errors within sample pixels of our no change strata. A larger sample might have revealed such limitations.

The themes measured by bare ground gain represent investment in energy, housing, and transportation, among other activities. As such, bare ground gain could be used as a possible indicator of economic activity. However, there are many challenges associated with this possibility

and further investigation would be required to establish credibility. For example, energy extraction is not uniform in footprint and may even occur offshore. Patterns of urbanization are similarly not uniform in space with factors including planning and zoning, settlement population density, and geographic constraints such as water bodies and terrain that can alter area of change for a given investment. Infrastructure investments are similarly not equally manifested on the land surface. The test of the utility of such an approach will be to relate it to existing economic data despite these mapping caveats.

Results indicate global and regional bare ground gain for areas experiencing rapid economic and demographic change. Seto et al. (2012) recognized urban population and Gross Domestic Product (GDP) as key determinants of urban expansion and developed a model to forecast global urban expansion to 2030. Our results support this narrative except for Africa and South Asia, two regions with rapid population growth, but comparatively little documented bare ground gain. Among the top ten countries with the largest bare ground gain estimates, China had a similar GDP growth rate with the U.S. before 2007. China also had the highest rate of urban population growth. India, on the other hand, had high urban population increase, but much slower GDP growth (World Bank, 2016b). This indicates the complex relationships of urban population and GDP change on human-induced bare ground gain. In-depth investigation of bare ground gain area and analysis of socioeconomic drivers on regional and country scales is important for scientists, planners and policy makers to understand the drivers and resultant rates of this most intensive of land cover conversions.

Acknowledgements

Support for this research was provided by NASA Land Use Land Cover Change Program (NNX15AK65G), Gordon and Betty Moore Foundation (GBMF-5131) and USGS Global Land Cover Initiative (G12AC20041). The authors thank Alexey Egorov from South Dakota State University who provided the WELD bare ground product. We also thank Viviana Zalles, Chima Okpa and Ricardo Aguilar from UMD GLAD team for assisting with training and validation interpretation. We appreciate valuable comments by three anonymous reviewers.

Appendix. Accuracy assessment formulas

The sampling design was stratified random sampling with the sampling unit defined as a Landsat pixel. A simple random sample was selected in each stratum h . The inclusion probability for pixel u in stratum h is $\pi_u = n_h/N_h$ with a stratum population size (N_h) and stratum sample size (n_h). The number of pixels in the population (i.e., region of interest) is N . Stehman (2014) derived the stratified estimates of accuracy using indicator functions (Cochran, 1977) y_u and x_u , where these observations obtained for pixel u have just two possible values, 0 or 1. For estimating user's accuracy of bare ground gain, define $y_u = 1$ if sample pixel u is correctly classified as bare ground gain, otherwise $y_u = 0$, and define $x_u = 1$ if sample pixel u is classified as bare ground gain, otherwise $x_u = 0$. For the producer's accuracy of bare ground gain, define $y_u = 1$ if sample pixel u is correctly classified as bare ground gain, otherwise $y_u = 0$, and define $x_u = 1$ if sample pixel u has reference class of bare ground gain, otherwise $x_u = 0$. The estimator of user's accuracy and producer's accuracy is then expressed as a ratio estimator:

$$\hat{R} = \frac{\sum_{h=1}^{25} N_h \bar{y}_h}{\sum_{h=1}^{25} N_h \bar{x}_h} \quad (A1)$$

648 where h is the stratum index, $\bar{y}_h = \sum_{u \in h} y_u / n_h$ is the sample mean of the y_u values in stratum h ,
 649 and $\bar{x}_h$ is the sample mean of the x_u values of stratum h . The variance estimator for $\hat{R}$ is:

$$650 \quad \hat{V}(\hat{R}) = \frac{1}{\hat{R}^2} \sum_{h=1}^{25} N_h^2 (1 - n_h/N_h) (s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R} s_{xyh}) / n_h \quad (A2)$$

651 where $\hat{R} = \sum_{h=1}^{25} N_h \bar{x}_h$, s_{yh}^2 is the sample variance of the y_u values in stratum h , s_{xh}^2 is the sample
 652 variance of the x_u values within stratum h , and s_{xyh} is the sample covariance between x_u and y_u
 653 of stratum h (see equations A3-A5 below):

$$654 \quad s_{yh}^2 = \sum_{u \in h} (y_u - \bar{y}_h)^2 / (n_h - 1) \quad (A3)$$

$$655 \quad s_{xh}^2 = \sum_{u \in h} (x_u - \bar{x}_h)^2 / (n_h - 1) \quad (A4)$$

$$656 \quad s_{xyh} = \sum_{u \in h} (y_u - \bar{y}_h)(x_u - \bar{x}_h) / (n_h - 1) \quad (A5)$$

657 To estimate overall accuracy, define $y_u = 1$ if pixel u is classified correctly and $y_u = 0$ if pixel u
 658 is classified incorrectly. The estimator for overall accuracy is then expressed as:

$$659 \quad \hat{O} = \sum_{h=1}^{25} N_h \bar{y}_h / N \quad (A6)$$

660 The variance estimator for $\hat{O}$ is:

$$661 \quad \hat{V}(\hat{O}) = \frac{1}{N^2} \sum_{h=1}^{25} N_h^2 (1 - n_h/N_h) s_{yh}^2 / n_h \quad (A7)$$

662 where s_{yh}^2 is computed as shown by equation (A3).

663 References

- 664 Alberti, M., 2005. The Effects of Urban Patterns on Ecosystem Function. *Int. Reg. Sci. Rev.* 28,
 665 168–192. doi:10.1177/0160017605275160
 666 Anderson, J.R., 1976. A land use and land cover classification system for use with remote sensor

667 data. US Government Printing Office.

668 Bonan, G.B., Pollard, D., Thompson, S.L., 1992. Effects of Boreal Forest Vegetation on Global
669 Climate. *Nature* 359, 716–718. doi:10.1038/359716a0

670 Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.

671 Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees.
672 CRC press.

673 Broich, M., Stehman, S. V, Hansen, M.C., Potapov, P., Shimabukuro, Y.E., 2009. A comparison
674 of sampling designs for estimating deforestation from Landsat imagery: A case study of the
675 Brazilian Legal Amazon. *Remote Sens. Environ.* 113, 2448–2454.

676 Cochran, W.G., 1977. Sampling techniques, 3rd ed. John Wiley & Sons, New York.

677 Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Review Article Digital
678 change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.* 25,
679 1565–1596. doi:10.1080/0143116031000101675

680 DeFries, R.S., Rudel, T., Uriarte, M., Hansen, M., 2010. Deforestation driven by urban
681 population growth and agricultural trade in the twenty-first century. *Nat. Geosci.* 3, 178–
682 181. doi:10.1038/ngeo756

683 Ellis, E.C., Ramankutty, N., 2008. Putting people in the map: anthropogenic biomes of the
684 world. *Front. Ecol. Environ.* 6, 439–447. doi:10.1890/070062

685 Foley, J.A., 2005. Global Consequences of Land Use. *Science* 309, 570–574.
686 doi:10.1126/science.1111772

687 Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H.,
688 Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C.,
689 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens.*

690 Environ. 83, 287–302.
 691 Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F.,
 692 Obersteiner, M., 2009. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land
 693 Cover. Remote Sens. 1, 345–354. doi:10.3390/rs1030345
 694 Geist, H.J., Lambin, E.F., 2002. Proximate causes and underlying driving forces of tropical
 695 deforestation. Bioscience 52, 143–150. doi:10.1641/0006-
 696 3568(2002)052[0143:pcaudf]2.0.co;2
 697 Gong, P., Wang, J., Yu, L., Zhao, Y.C., Zhao, Y.Y., Liang, L., Niu, Z.G., Huang, X.M., Fu,
 698 H.H., Liu, S., Li, C.C., Li, X.Y., Fu, W., Liu, C.X., Xu, Y., Wang, X.Y., Cheng, Q., Hu,
 699 L.Y., Yao, W.B., Zhang, H., Zhu, P., Zhao, Z.Y., Zhang, H.Y., Zheng, Y.M., Ji, L.Y.,
 700 Zhang, Y.W., Chen, H., Yan, A., Guo, J.H., Wang, L., Liu, X.J., Shi, T.T., Zhu, M.H.,
 701 Chen, Y.L., Yang, G.W., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z.L., Chen, J., 2013.
 702 Finer resolution observation and monitoring of global land cover: first mapping results with
 703 Landsat TM and ETM+ data. Int. J. Remote Sens. 34, 2607–2654.
 704 doi:10.1080/01431161.2012.748992
 705 Grimm, N.B., Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X.,
 706 Briggs, J.M., Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X.,
 707 Briggs, J.M., 2008. Global change and the ecology of cities. Science 319, 756–760.
 708 doi:10.1126/science.1150195
 709 Hansen, M.C., 2012. Classification trees and mixed pixel training data. Remote Sens. L. Use L.
 710 Cover Princ. Appl. 127–136.
 711 Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., Sohlberg, R.A.,
 712 2003. Global percent tree cover at a spatial resolution of 500 meters: First results of the

713 MODIS vegetation continuous fields algorithm. *Earth Interact.* 7, 1–15.

714 Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover
715 classification at 1km spatial resolution using a classification tree approach. *Int. J. Remote*
716 *Sens.* 21, 1331–1364. doi:10.1080/014311600210209

717 Hansen, M.C., Egorov, A., Potapov, P. V, Stehman, S. V, Tyukavina, A., Turubanova, S.A.,
718 Roy, D.P., Goetz, S.J., Loveland, T.R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C.,
719 Bents, T., 2014. Monitoring conterminous United States (CONUS) land cover change with
720 Web-Enabled Landsat Data (WELD). *Remote Sens. Environ.* 140, 466–484.
721 doi:10.1016/j.rse.2013.08.014

722 Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J.C., Turubanova, S., Kommareddy, I.,
723 Loveland, T.R., 2011. Continuous fields of land cover for the conterminous United States
724 using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project.
725 *Remote Sens. Lett.* 2, 279–288. doi:10.1080/01431161.2010.519002

726 Hansen, M.C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau,
727 D., Stehman, S. V, Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L.,
728 Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century
729 Forest Cover Change. *Science* 342, 850–853.

730 Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O., Altstatt, A., 2008. A method
731 for integrating MODIS and Landsat data for systematic monitoring of forest cover and
732 change in the Congo Basin. *Remote Sens. Environ.* 112, 2495–2513.
733 doi:10.1016/j.rse.2007.11.012

734 Homer, C., Huang, C.Q., Yang, L.M., Wylie, B., Coan, M., 2004. Development of a 2001
735 National Land-Cover Database for the United States. *Photogramm. Eng. Remote Sensing*

736 70, 829–840.

737 IPCC, 2015. Climate change 2014: mitigation of climate change. Cambridge University Press,
738 Cambridge, United Kingdom and New York, NY, USA.

739 Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change
740 detection method for updating the National Land Cover Database to circa 2011. Remote
741 Sens. Environ. 132, 159–175.

742 Kalnay, E., Cai, M., 2003. Impact of urbanization and land-use change on climate. Nature 423,
743 528–531. doi:10.1038/nature01675

744 Khorram, S., 1999. Accuracy Assessment of Remote Sensing-derived Change Detection.
745 American Society for Photogrammetry and Remote Sensing, Bethesda, Md.

746 Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM +
747 land surface observations and implications for global 30 m Landsat data product generation.
748 Remote Sens. Environ. 130, 280–293. doi:10.1016/j.rse.2012.12.003

749 Lambin, E.F., Geist, H.J., Lepers, E., 2003. Dynamics of land-use and land-cover change in
750 tropical regions. Annu. Rev. Environ. Resour. 28, 205–241.
751 doi:10.1146/annurev.energy.28.050302.105459

752 Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000.
753 Development of a global land cover characteristics database and IGBP DISCover from 1
754 km AVHRR data. Int. J. Remote Sens. 21, 1303–1330. doi:10.1080/014311600210191

755 Moran, M.S., Rahman, A.F., Washburne, J.C., Goodrich, D.C., Weltz, M.A., Kustas, W.P., 1996.
756 Combining the Penman-Monteith equation with measurements of surface temperature and
757 reflectance to estimate evaporation rates of semiarid grassland. Agric. For. Meteorol. 80,
758 87–109. doi:10.1016/0168-1923(95)02292-9

759 Olofsson, P., Foody, G.M., Herold, M., Stehman, S. V, Woodcock, C.E., Wulder, M.A., 2014.
 760 Good practices for estimating area and assessing accuracy of land change. *Remote Sens.*
 761 *Environ.* 148, 42–57. doi:http://dx.doi.org/10.1016/j.rse.2014.02.015
 762 Olofsson, P., Foody, G.M., Stehman, S. V, Woodcock, C.E., 2013. Making better use of
 763 accuracy data in land change studies: Estimating accuracy and area and quantifying
 764 uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 122–131.
 765 doi:http://dx.doi.org/10.1016/j.rse.2012.10.031
 766 Potapov, P. V., Turubanova, S.A., Tyukavina, A., Krylov, A.M., McCarty, J.L., Radeloff, V.C.,
 767 Hansen, M.C., 2015. Eastern Europe’s forest cover dynamics from 1985 to 2012 quantified
 768 from the full Landsat archive. *Remote Sens. Environ.* 159, 28–43.
 769 doi:10.1016/j.rse.2014.11.027
 770 Potapov, P. V, Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Mane, L.,
 771 Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo,
 772 2000-2010, with Landsat ETM plus data. *Remote Sens. Environ.* 122, 106–116.
 773 doi:10.1016/j.rse.2011.08.027
 774 Potere, D., 2008. Horizontal Positional Accuracy of Google Earth’s High-Resolution Imagery
 775 Archive. *Sensors* 8, 7973–7981. doi:10.3390/s8127973
 776 Schneider, A., 2012. Monitoring land cover change in urban and pen-urban areas using dense
 777 time stacks of Landsat satellite data and a data mining approach. *Remote Sens. Environ.*
 778 124, 689–704. doi:10.1016/j.rse.2012.06.006
 779 Schneider, A., Friedl, M.A., Potere, D., 2010. Mapping global urban areas using MODIS 500-m
 780 data: New methods and datasets based on “urban ecoregions.” *Remote Sens. Environ.* 114,
 781 1733–1746. doi:10.1016/j.rse.2010.03.003

782 Schneider, A., Friedl, M.A., Potere, D., 2009. A new map of global urban extent from MODIS
783 satellite data. *Environ. Res. Lett.* 4. doi:10.1088/1748-9326/4/4/044003

784 Serra, P., Pons, X., Sauri, D., 2003. Post-classification change detection with data from different
785 sensors: some accuracy considerations. *Int. J. Remote Sens.* 24, 3311–3340.

786 Seto, K.C., Guneralp, B., Hutyrá, L.R., 2012. Global forecasts of urban expansion to 2030 and
787 direct impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16083–
788 16088. doi:10.1073/pnas.1211658109

789 Sexton, J.O., Song, X.-P., Huang, C., Channan, S., Baker, M.E., Townshend, J.R., 2013. Urban
790 growth of the Washington, DC-Baltimore, MD metropolitan region from 1984 to 2010 by
791 annual, Landsat-based estimates of impervious cover. *Remote Sens. Environ.* 129, 42–53.
792 doi:10.1016/j.rse.2012.10.025

793 Shukla, J., Nobre, C., Sellers, P., 1990. Amazon deforestation and climate change. *Science* 247,
794 1322–1325. doi:10.1126/science.247.4948.1322

795 Simons, H., Soto, X., Zhu, Z., Singh, K.D., Bellan, M.-F., Iremonger, S., Hirvonen, H., Smith,
796 B., Watson, V., Tosi, J., 2001. Global Ecological Zoning for the Global Forest Resources
797 Assessment 2000-Final Report.

798 Singh, A., 1989. Review Article Digital change detection techniques using remotely-sensed data.
799 *Int. J. Remote Sens.* 10, 989–1003. doi:10.1080/01431168908903939

800 Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-
801 OLS night lights. *Remote Sens. Environ.* 96, 277–291. doi:10.1016/j.rse.2005.02.002

802 Song, X.-P., Sexton, J.O., Huang, C., Channan, S., Townshend, J.R., 2016. Characterizing the
803 magnitude, timing and duration of urban growth from time series of Landsat-based
804 estimates of impervious cover. *Remote Sens. Environ.* 175, 1–13.

doi:10.1016/j.rse.2015.12.027

Stehman, S. V., 2013. Estimating area from an accuracy assessment error matrix. *Remote Sens. Environ.* 132, 202–211. doi:10.1016/j.rse.2013.01.016

Stehman, S. V, 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *Int. J. Remote Sens.* 35, 4923–4939. doi:10.1080/01431161.2014.930207

Tyukavina, A., Baccini, A., Hansen, M.C., Potapov, P. V, Stehman, S. V, Houghton, R.A., Krylov, A.M., Turubanova, S., Goetz, S.J., 2015. Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. *Environ. Res. Lett.* 10, 74002. doi:10.1088/1748-9326/10/7/074002

Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of Earth's ecosystems. *Science* 277, 494–499. doi:10.1126/science.277.5325.494

Wang, B., Huang, J., Yang, X., Zhang, B., Liu, M., 2010. Estimation of biomass, net primary production and net ecosystem production of China's forests based on the 1999-2003 National Forest Inventory. *Scand. J. For. Res.* 25. doi:10.1080/02827581.2010.524891

Wang, L., Li, C., Ying, Q., Cheng, X., Wang, X., Li, X., Hu, L., Liang, L., Yu, L., Huang, H., 2012. China's urban expansion from 1990 to 2010 determined with satellite remote sensing. *Chinese Sci. Bull.* 57, 2802–2812.

Weber, K.T., Alados, C.L., Guillermo Bueno, C., Gokhale, B., Komac, B., Pueyo, Y., 2009. Modeling Bare Ground With Classification Trees in Northern Spain. *Rangel. Ecol. Manag.* 62, 452–459.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail,

828 P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., Landsat Sci, T., 2008. Free
829 access to Landsat imagery. *Science* 320, 1011.

830 World Bank, n.d. World Bank Country and Lending Groups [WWW Document]. URL
831 [https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups)
832 [and-lending-groups](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups) (accessed 10.7.16a).

833 World Bank, n.d. World Development Indicators-GDP growth [WWW Document]. URL
834 <http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG> (accessed 10.7.16b).

835 Xian, G., Homer, C., 2010. Updating the 2001 National Land Cover Database Impervious
836 Surface Products to 2006 using Landsat Imagery Change Detection Methods. *Remote Sens.*
837 *Environ.* 114, 1676–1686. doi:10.1016/j.rse.2010.02.018

838 Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., Wickham, J., 2011. The change of
839 impervious surface area between 2001 and 2006 in the conterminous United States.
840 *Photogramm. Eng. Remote Sensing* 77, 759–762.

841 Yang, L., Huang, C., Homer, C.G., Wylie, B.K., Coan, M.J., 2003. An approach for mapping
842 large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial
843 resolution imagery. *Can. J. Remote Sens.* 29, 230–240. doi:10.5589/m02-098

844 Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically
845 mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24, 583–594.
846 doi:10.1080/01431160210144570

847 Zhan, X., Sohlberg, R.A., Townshend, J.R.G., DiMiceli, C., Carroll, M.L., Eastman, J.C.,
848 Hansen, M.C., DeFries, R.S., 2002. Detection of land cover changes using MODIS 250 m
849 data. *Remote Sens. Environ.* 83, 336–350. doi:10.1016/s0034-4257(02)00081-0

850 Zhang, Q., Seto, K.C., 2011. Mapping urbanization dynamics at regional and global scales using

multi-temporal DMSP/OLS nighttime light data. Remote Sens. Environ. 115, 2320–2329.
doi:10.1016/j.rse.2011.04.032

LIST OF FIGURE CAPTIONS

Fig.1. Work flow of global bare ground gain characterization. Annual Landsat mosaics are inputs. Two bare ground gain characterization methods are shown: direct bare ground mapping and indirect bare ground gain mapping. Colored boxes identify training data, algorithm and output maps applied to each mapping method. Orange boxes at bottom show the structure of mapped strata for validation and sample-based area estimation. Note that mapping is not used for area estimation, but as strata for sample-based area estimation.

Fig.2. Band profiles of bare ground gain pixels from (a) a coal mining site in Alabama, USA, centered at 33°53'20"N, 87°32'36"W and (b) a power plant in North Carolina, USA, centered at 34°50'32"N, 79°44'8"W, above which are Landsat false color composites of SWIR (1.6 μ m), NIR and RED bands corresponding to the start, change and end year.

Fig.3. Example of stratification for Brawley, California (a 10 km by 6.5 km extent centered at 32°59'5"N, 115°30'33"W), within the subtropical domain. Strata are shown on the right image where red is jointly mapped gain (Stratum 6), yellow is directly mapped (Stratum 7), blue is indirectly mapped (Stratum 8), and dark gray is the 90m buffer (Stratum 9); white is mapped no gain (Stratum 10).

Fig.4. Stratified random sample with 25 strata defined from the cross classification of five climate domains with five map feature classes.

872 Fig.5. Boundaries of geographic regions from the World Bank for regional bare ground gain area
873 reporting. The sample size located in each region is: East Asia and Pacific 1444, Europe and
874 Central Asia 1416, North America 987, Latin America and the Caribbean 608, South Asia 486 and
875 Sub-Saharan Africa 411, Middle East and North Africa 377. There are 21 sample pixels falling
876 outside of the boundaries that we excluded from area estimation.

877 Fig.6. Landsat composites and directly classified bare ground gain: (a) Petroleum drilling in rural
878 Texas centered at $30^{\circ}24'33''\text{N}$, $101^{\circ}02'46''\text{W}$; (b) Urbanization in Beijing, China centered at
879 $39^{\circ}48'01''\text{N}$, $116^{\circ}33'07''\text{E}$. The first and last composites are Landsat false color composites of
880 SWIR ($1.6\ \mu\text{m}$), NIR and RED bands. Bare ground gain is shown in red on images of right panel
881 with background displayed in the false color composite of median_max, median_latest and
882 median_earliest in percent bare ground cover as RGB.

883 Fig.7. Percent bare ground gain aggregated from Landsat strata to 0.05° resolution in lat/lon
884 geographic projection.

885 Fig.8. Examples of characterized bare ground gain land cover / land use outcomes in a $930\text{m} \times$
886 930m (31×31 Landsat pixels) block centered at sample points as shown in red square: (a)
887 Resource extraction, coal mine in Queensland, Australia centered at $21^{\circ}29'29''\text{S}$ $148^{\circ}23'12''\text{E}$; (b)
888 Infrastructure development, airport runway in Chongqing, China centered at $29^{\circ}42'29''\text{N}$
889 $106^{\circ}38'35''\text{E}$; (c) Commercial/residential built-up, shopping plaza in Oregon, U.S. centered at
890 $45^{\circ}0'46''\text{N}$ $122^{\circ}59'54''\text{W}$; (d) Transitional bare ground gain, bare field next to a building in Rio
891 Grande do Sul, Brazil centered at $29^{\circ}51'7''\text{S}$ $50^{\circ}14'53''\text{W}$; (e) Natural bare ground gain, river
892 meander in Assam, India centered at $27^{\circ}46'45''\text{N}$ $95^{\circ}39'11''\text{E}$; (f) Greenhouses in Antalya, Turkey
893 centered at $37^{\circ}0'16''\text{N}$ $30^{\circ}50'49''\text{E}$. Images from left to right column are: 1), 2) reference Google

894 Earth time one and time two imagery; 3),4) Landsat start and end date composites of SWIR (1.6
895 μm), NIR and RED bands as RGB; 5) Direct bare ground gain classification results with gain in
896 yellow and no gain in black.

897 Fig. 9. Bare ground gain area estimates for 2000-2012 (95% confidence intervals represented by
898 error bars) and composition of land cover / land use outcomes of bare ground gain: a) regional
899 estimates; b) estimates of top ten countries with the largest directly mapped bare ground gain area.

900 Fig.10. Proportions of estimated bare ground gain area to land cover / land use outcomes: a) by
901 region and b) by country.