**Identifying agricultural frontiers for modelling global cropland expansion**

Felix Eigenbrod1\*, Michael Beckmann2, Sebastian Dunnett1, Laura Graham1, Robert A. Holland1, Patrick Meyfroidt3,4, Ralf Seppelt2,5,6, Xiao-Peng Song7, Rebecca Spake1, Tomáš Václavík8,9, Peter H. Verburg10,11

1. School of Geography and Environmental Science, University of Southampton, UK
2. Department of Computational Landscape Ecology, UFZ—Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany
3. Earth and Life Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium
4. Fonds de la Recherche Scientifique (F.R.S.- FNRS), 1000 Brussels, Belgium
5. iDiv—German Centre for Integrative Biodiversity Research, 04103 Leipzig, Germany.
6. Institute of Geoscience & Geography, Martin-Luther-University Halle-Wittenberg, 06099 Halle (Saale), Germany
7. Department of Geosciences, Texas Tech University, Lubbock, TX 79409, USA
8. Department of Ecology and Environmental Sciences, Faculty of Science, Palacký University Olomouc, 78371 Olomouc, Czech Republic
9. Global Change Research Institute of the Czech Academy of Sciences, 60300 Brno, Czech Republic
10. Institute for Environmental Studies, VU University Amsterdam, de Boelelaan 1087, 1081HV Amsterdam, the Netherlands
11. Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland

\* Lead contact: [f.eigenbrod@soton.ac.uk](mailto:f.eigenbrod@soton.ac.uk)

**Abstract**

The increasing expansion of cropland is major driver of global carbon emissions and biodiversity loss. However, predicting plausible future global distributions of croplands remains challenging. Here, we show that, in general, existing global data aligned with classical economic theories of expansion explain the current (1992) global extent of cropland reasonably well, but not recent expansion (1992-2015). Deviations from models of cropland extent in 1992 (‘frontierness’) can be used to improve global models of recent expansion, likely as these deviations are a proxy for cropland expansion under frontier conditions where classical economic theories of expansion are less applicable. Frontierness is insensitive to the land cover dataset used, and is particularly effective in improving models that include mosaic land cover classes and the largely smallholder driven frontier expansion occurring in such areas. Our findings have important implications as the frontierness approach offers a straightforward way to improve global land use change models.

**Introduction**

The increasing appropriation of land by humans is a major driver of global environmental change 1,2. Cropland expansion is a particular issue, given the projected increase in demand for agricultural products driven by changing dietary patterns of a growing population, increased demand for land-based climate solutions3, and the potential threat such expansion has for high-biodiversity tropical areas 4–6. While intensification of agriculture on existing agricultural lands can reduce the need for expansion into uncultivated areas 7,8, the rate of global yield gains per area is decreasing 9, and the rate of agricultural expansion has been increasing since 2000, mostly in the tropics 10. However, much uncertainty exists about how much future expansion of cropland is likely to occur globally 11, and where this will occur 12–14. This creates uncertainty in cropland expansion projections, which in turn has major implications for global models of land-based climate change mitigation options and of biodiversity loss 11,12.

A key challenge to forecasting cropland expansion is the uncertainty as to whether the identity and magnitude of predictors of cropland expansion are consistent across space and time. While such consistency may be true in consolidated agricultural regions such as Europe 15, this is unlikely to be true globally, because frontier expansion of agriculture occurs under very different circumstances than in consolidated agricultural regions 16,17. Frontiers are regions with rapid land-use expansion and an imbalance between abundant land and natural resources and a relative lack of capital or labour to exploit these resources16,18. In frontier areas, case studies and theory suggest that agricultural expansion occurs under imperfect market conditions where agency of key actors, rent-capture behaviours and non-equilibrium dynamics dominate 18*.* Large-scale actors operating in commodity frontiers have the agency to modify the very conditions (such as accessibility, infrastructures, policies) that influence their own cropland expansion 18. By contrast, in smallholder-driven frontiers, coexistence between market-oriented and subsistence production and imperfect market integration are such that decisions may differ from the simple economic optimum captured in coarse-scale global approaches19,20. Finally, the globalized nature of trade in land-based products such as agricultural commodities and timber 21 may lead to amplification of regional disparities in land use trajectories, further increasing global differences between past and current predictors of cropland expansion.

Frontier dynamics remain poorly represented in global land use models 13 as even the most recent models are largely based on classical land rent theories under equilibrium conditions 20,22. Moreover, systematic tests of the effects of these assumptions on the accuracy of the predictions of global land use models remain limited 23, and the spatial and temporal differences in the identity and effects of putative predictors of agricultural expansion have yet to be tested at the global scale. This lack of testing represents a key research gap for predicting future land cover changes 24.

Until recently, a major barrier to assessing whether predictors of land use change are stable or changing over time has been the lack of finely resolved, global data on land use change. Such data are now available, at least for recent changes in forest cover 25, and land cover (1992-2015) 26. A recent global analysis suggested that forest loss between 2000 and 2015 was driven both by commercial agriculture and shifting (subsistence) agriculture 27, but did not directly assess how this relates to recent agricultural expansion, nor whether predictors of forest loss are stationary over time. Therefore, the need remains for an improved global, spatially resolved understanding of which socio-ecological factors are the most important predictors of recent agricultural expansion 28, and if and how these differ from the factors that explain the current global extent of cropland.

Here, we address this challenge in two ways. First, we use logistic regression models to compare how well globally available, independent putative predictors of agriculture explain global cropland **extent** (in 1992) and recent cropland **expansion** (1992-2015). This provides a first global test of the degree to which the consistency and magnitude of predictors of cropland remain consistent over time. Globally available data enables tests of classical theories of cropland expansion but not of theories of recent expansion in agricultural frontiers 18,20 (Table 1), so our *a priori* hypothesis was that this data would better predict extent than expansion of cropland, as the latter is likely mostly occurring in frontier regions 20.

Second, we applied positive deviance analysis 29 to the extent of cropland in 1992 to develop and test the utility of a new proxy for recent agricultural expansion frontiers (which we refer to hereon as ‘**frontierness**’). Positive deviance (or ‘bright spot’) analysis has been used in fields including human development 29, conservation 30 and ecosystem services 31, to identify locations where the outcome variable exceeds (or falls below, cf. ‘dark spots’) expectations derived from a null model. The null model should be clearly linked to hypotheses about known factors 31. Here, the ‘**Null Model**’ corresponds to the classical theories of agricultural expansion that can be quantified globally with available data (Table 1) (Experimental Procedures). Positive deviations from the Null Model are therefore locations with cropland in 1992 *not* predicted by classical theories of agricultural expansion, and as such potentially indicative of pre-1992 cropland expansion frontiers. We then tested if these deviations from the Null Model of extent in 1992 – frontierness – can explain expansion of cropland from 1992 to 2015. The hypothesis we test here is that pre-1992 frontiers are correlated with short-to-medium term post-1992 cropland expansion frontiers.

We show that, in general, existing global data, which are aligned with classical theories of agricultural expansion, explain the current global extent of cropland reasonably well, but do not explain most recent cropland expansion well. However, our proxy for frontiers (frontierness) generally provides significant improvements to models of 1992-2015 cropland expansion. Our findings have important implications for global land use change models, as they show both that the dependence of these models of classical theories of agricultural expansion is problematic, but also that they can be improved using the frontierness approach we demonstrate here.

**Results**

*Standard theories explain extent better than expansion*

In general, independent global predictor variables aligned with classical theories of cropland expansion (Table 1; Fig 1 & Model 1; Experimental Procedures) explained the 1992 extent of cropland globally well, but not recent expansion (1992-2015). However, the choice of threshold used to categorise land into non-cropland and cropland (to perform logistic regressions) had a major effect on the expansion analyses, though not the extent analyses (Fig. 1). Expansion events were characterised as extreme (that is, very high prevalence of within-pixel expansion) (>50% of a 5’ x 5’ pixel, subsequently expansion50%), major (>10% of a 5’ x 5’ pixel, subsequently expansion10%), or minor (>0.5% of a 5’ x 5’ pixel, subsequently expansion0.5%). Extreme within-pixel expansionevents are relatively predictable using existing global datasets (average McFaddden pseudo-*R*2 expansion50% = 0.28), but these events are globally very rare (only 0.45% of expansion events at the 0.5% threshold are also above the 50% threshold; Fig. S1 in Supplementary Information). However, most major and minor recent expansion is poorly predicted (pseudo-*R*2 expansion10% = 0.12; expansion0.5% = 0.08) using this existing data. By contrast, results for extent were relatively consistent, irrespective of the threshold chosen (average McFadden pseudo-*R*2: extent0.5% = 0.32, extent10% = 0.30, extent50% = 0.33).

The magnitude and even direction of the effects of some predictors varied across models of cropland extent and recent expansion. The effect of GDP (a proxy for economic activity) was positive for the extent models, but negative for the expansion models. At a 50% threshold, ‘Access’ (distance to markets) had a positive relationship with cropland for the extent model, but a negative effect on expansion. The considerably higher explanatory power of the 50% threshold expansion models relative to the lower thresholds appears to be due to the much stronger negative effect of ‘steepness’ (percentage cover of steep slopes), and to a lesser degree the positive effect of population density for this threshold relative to expansion at other thresholds. Differences between thresholds for the extent models were less pronounced, but steepness also had a higher explanatory power for extent of agriculture in 1992 for the 50% threshold than the other two thresholds (Fig. 1). See Fig S2 in the SI for global distributions of extent and expansion of cropland at all three thresholds.

*Expansion aligns with positive deviance*

A disproportionate amount of recent expansion of cropland (0.5% or 10% threshold) occurs in areas that have at least 1 SD or more cropland in 1992 (positive deviations) than would be predicted by the predictors of classical theory within the Null Model (Table 2; Fig. 1). Such areas of high positive deviations correspond to high values of frontierness (all deviations from the 1992 Null Model; Null Model coefficients in Table S1). Recent expansion at the 0.5% threshold is extremely common (21% of bioclimatically suitable areas), including within regions already dominated by agriculture (Europe, eastern USA, China), as well as within frontier regions in South America, sub-Saharan Africa and the former USSR. Overlaps with positive deviations of 1 or more SD is high within the latter areas – with the exception of Indonesia – but not the former. Major expansion (10% threshold) is much less common, and concentrated in frontiers regions, where it shows a high degree of overlap with positive deviations from the Null Model (Fig 2). Conversely, very rare, extreme expansion (50% threshold) occurs in areas with less cropland in 1992 (negative deviations) than would be predicted by classical theory (Table 2).

*Frontierness improves models of expansion*

Inclusion of frontierness as a predictor (Model 3) increased the predictive power of the 0.5% and 10% threshold 1992-2015 expansion models by a pseudo-*R*2 of 0.08 over models using existing predictors (Model 2), but did not improve the predictive power of the 50% threshold expansion model (Table 3).

**Discussion**

A key challenge for predicting future land use change globally is understanding the degree to which the predictors of such changes are stationary in space and time 12,28. Here we show that while classical theories of cropland expansion explain the current (1992) extent of cropland reasonably well, this is generally not the case for recent cropland expansion (1992-2015). We also show that using deviations from a null model of cropland extent in 1992 (frontierness) improves our ability to predict recent expansion. Both results have important implications for global land use models.

At present, approaches to modelling gridded global land use and land cover change generally use existing cropland and available land as initial inputs for prediction, in addition to measures of regional disaggregated demand for agricultural products 12. Some gridded models also assume expansion after accounting for constraints (e.g. protected area coverage, bioclimatic limits of cropland, other existing land cover types) and endogenous suitability such as soil type and aridity 24,32. Our findings show this approach is unlikely to predict future agricultural expansion well, most likely because most current cropland expansion is driven by multiple processes and actors that are not fully captured by existing global datasets 17,18. Key commercial crops do disproportionately overlap with cropland expansion at all thresholds (Table S8); however, such expansion can be driven both by large commercial land holders33 and smallholders 34. More generally, the disproportionate overlap of expansion at the 0.5% and 10% threshold with both large and small fields (Table S8) supports existing work showing that frontier expansion is driven both by commercial and subsistence activities 16–18,35.

Our proxy for agricultural expansion frontiers – frontierness – added explanatory power to models of agricultural expansion, except for the most extreme, and very rare, expansion events (>50% threshold). The spatial pattern of our frontierness indicator appears to align with recent cropland expansion in areas recognized as key global cropland expansion frontiers in South America (Cerrado and Chaco ecoregions 33), as well as forest frontiers in West Africa, and a large arc of cropland expansion in the steppes of Southern Russia and Kazakhstan – some of these the result of recultivation of land abandoned at the collapse of the Soviet Union 36–38 (Fig 3). This is likely because the post-1992 expansion frontiers occur largely in the same places as recent, pre-1992 expansion frontiers. Indeed, post-hoc analyses showed proportionally similar or higher overlap with areas of high positive deviation from the Null Models of cropland extent for agriculture-driven deforestation in the period 1982-1992 as compared to 1992-2015, as measured using an independent global dataset 39 (Supplementary Information; Supplementary Experimental Procedures and Table S9). This suggests that at present there is at least some medium term (~30 years) stationarity in the bulk of global cropland expansion frontiers. This is likely because expansion frontiers build on agglomeration economies, which require some time to appear as cropland expands in a region, and thus cropland expansion disproportionately occurs near places where some expansion has already taken place 20,40. As such, at the 5’ x 5’ arc-min resolution (roughly 10 x 10 km at the equator), pixels can remain ‘expansion’ pixels over a fairly long time period, as gradually more of their area is converted to cropland. Interestingly, *post-hoc* overlap analyses (Table S2) demonstrated that areas with more cropland than predicted (high positive deviations from the Null Models of cropland extent in 1992) disproportionately overlapped with the largest fields, as measured in 201741. This might reflect the recent shifts to large-scale, commercial agriculture in frontier areas, as observed in South America33, and that recent expansion is predominantly associated with large-scale actors more generally. This post-hoc analysis also showed no tendency of positive deviations towards less developed countries (as measured by political stability or HDI), and only a weak positive effect of favourable soil conditions, again illustrating the difficulty in a priori identification of expansion frontiers using existing datasets. Overlaps of these variables with areas with high negative deviations from the predicted extent of cropland in 1992 (so less cropland than expected) did appear to be disproportionately located in areas with major soil constraints; this perhaps explains the overlaps with small field sizes in such areas. As the frontierness proxy is all deviations from the Null Model (positive and negative), it suggests that negative values of the proxy flag areas where local conditions (e.g. soil constraints) limit cropland expansion that is otherwise predictable from existing global datasets.

As current frontiers can be identified using our positive deviance frontierness approach for the most current available land cover data, this means we can use the approach to predict medium-term future frontier areas with currently available data within existing predictive models of land use change. Frontierness could therefore be used to help close a major research gap – building spatially dissagregated transfer functions within gridded economic models 28. For example, frontierness could be an additional input into global, country-level economic models, enabling the modelling of within-country spatial heterogeneity in land supply elasticities, building on recent work where market access was used within such models 42. Moreover, the same positive deviance approach that underpins frontierness could potentially be used for improving models of other types of land use or land-use intensity change. Of course, if the nature of expansion changes – for example to more rapid, large-scale expansion in frontier areas, as appears to be the case in South America 33 – the predictive power of frontierness may decrease in the future. However, even where regions are changing, frontierness could still be useful, as a way of helping to identify outlier regions, where additional case studies and ground-based analyses could help understand why they deviate from expectations and thereby enrich land use theories. For example, the relative *lack* of overlap of frontierness with recent major (>10%) expansion in Indonesia (Fig. 3) could be partially linked to the 2011 moratorium on new land concessions in Indonesia, which slowed deforestation since 2012 43. To encourage further use of our proxy, we provide global geospatial files of frontierness – calculated for both 1992 (as used in this paper) and for 2015 (latest data) as outlined in the Data and Code Availability in the Resource Availability section.

Counterintuitively, very rare, extreme expansion (50% threshold) can be relatively well explained by existing predictor variables. *Post-hoc* overlap analysis showed that such expansion has a disproportionate overlap with modelled distributions of soy and very large fields (Supplementary Information; Table S8). Large fields are often associated with large farms, which in turn relate to farm income 35,44,45. As such, observed extreme expansion fits with the commodity and neoliberal resource frontier theories of agricultural expansion. Here, rapid land use expansion for large-scale agriculture occurs in resource frontiers where there is an abundance of land on which structural factors such as new agricultural technologies, infrastructure, and rising producer prices, facilitate large agricultural rents, but the expansion strongly depends on the ability of powerful actors to capture or influence these rents through large-scale, capital-intensive agriculture (e.g.18,46). This might include influencing the location of new roads or infrastructures like ports, or even building these themselves, influencing the spatial pattern and implementation of land use policies such as areas and practices targeted for agricultural subsidies or where land uses are restricted 18. This might explain why these rare expansion events exhibit a negative, rather than positive response to distance to market and population density. Market access may also be acting as a proxy for some predictors of commodity frontiers as it is inversely correlated with poverty and low land costs 47. Our ability to explain these very rare, extreme expansion events reasonably well with global data may be also partly be due to a relatively small number of 5’ × 5’ pixels globally that are 1) flat enough for intensive mechanized agriculture (steepness is a key predictor of extreme expansion); 2) where some cropland already exists (agriculture in 1992 is a key predictor as well), but 3) there is still enough space for the ~25 km2 (at the equator) of new cropland this threshold entails. The relatively high predictability of rare, extreme cropland expansion events is an important finding, as while very rare, such events represent the extreme points of wider, industrial-scale expansion frontiers. Such areas (e.g. the South American resource frontiers, which are increasingly characterized by large field sizes 33) are known to have major social and ecological consequences 46,48.

Importantly, our ‘frontierness’ approach is robust to how cropland is defined, the choice and thematic resolution of the inputted cropland map, as well as the length of the time series considered. Results of including mosaic land cover classes (hereafter ‘mosaic’) were similar to the main analyses for the ‘extent’ analyses (Table S5). The explanatory power of existing predictors (Model 2) was lower in the ‘mosaic’ analyses for the 10% and 50% thresholds for the base ‘expansion’ models (0.16 vs 23 and 0.30 vs 0.41, respectively). However, importantly, for the models that included frontierness (Model 3), the explanatory power of the ‘mosaic’ models was identical to the main results (Table S6). These results suggest that existing predictors capture mosaic expansion poorly, likely as such mosaic expansion represents a smallholder driven expansion frontier. Fortunately, our frontierness proxy is effective in explaining such mosaic expansion frontiers when the Null Model includes mosaic land cover classes in estimates of cropland. The choice of end date for the ‘expansion’ analyses was unimportant; end dates of 2013 and 2015 yielded almost identical explanatory power of all ‘expansion’ models (Table S6). Finally, results of ‘extent’ analyses using either the ESA CCI data (main results) and MODIS data49 for 2001-2015 also differed negligibly, likely as both datasets accurately represent the majority of cropland at the global scale50(Supplementary Experimental Procedures; Figs S3 & S4; Table S7). However, the explanatory power of models using MODIS data was much higher than the CCI data for the 2001-2015 ‘expansion’ analyses; indeed, ‘expansion’ is better predicted than ‘extent’ based on the MODIS data (Table S7). Critically, adding ‘frontierness’ still greatly improved the explanatory power (pseudo R2) of the MODIS expansion models (from 0.35 to 0.53 and 0.42 to 0.57 for 0.5% and 10% cropland, respectively)(Table S7), again illustrating the robustness of our frontierness approach to varying definitions of cropland. The relative predictability of expansion as measured by MODIS is perhaps because underrepresentation of cropland by the MODIS data in 2001 (Fig S3) meant more of the measured expansion 2001-2015 occurred within relatively predictable consolidated agricultural regions as opposed to frontier regions (Fig S4).

As with any global analyses, our work here is subject to a number of caveats. Firstly, we were unable to address adequately in the ‘extent’ analysis the issue of endogeneity of human population density, GDP and cropland extent. Population growth is a known driver of cropland expansion51, but people are drawn to existing agricultural areas, as reflected by the establishment of many of the world’s great cities within areas with rich agricultural lands. As population density is measured in 1990, and cropland and GDP in 1992, it is impossible to ascertain statistically which ones are causal. The issue of endogeneity also affects the ‘market access’ layer, as this is based on city populations and human populations in 200052. Unfortunately, these issues of endogeneity cannot be addressed as no spatially resolved global primary datasets of human population exist prior to 1990; as such the ‘extent’ analyses should be viewed with more caution than the ‘expansion’ analyses, which do not suffer from this problem. Secondly, as gridded global time series for key potential drivers of expansion (e.g. population) do not exist prior to 1990, we were unable to test if past trajectories of change of population (or GDP) can improve the predictive power of models of cropland expansion and reduce the utility of frontierness. Changes in population density, human development index (HDI) and GDP between 1990 and 2015 (post-hoc overlap analyses; see Supplementary Experimental Procedures and Table S8) were not associated with cropland expansion at the 0.5% and 10% threshold. However, it is possible that longer-term, and lagged changes in socio-economic variables could be important predictors of future expansion; testing this should be a priority as such data becomes available.

Accurate predictions of cropland expansion are critical in understanding both issues of food security and climate change 53, but also how these in turn affect biodiversity 6,54. This study shows that in general recent cropland expansion cannot be explained well with currently available data. Our study also highlights the value using residual deviations from a model of current extent of cropland (our frontierness proxy) as a new way of improving the predictive power of models of cropland expansion over and above what is possible with globally available predictor variables. This is likely because these residuals, potentially representing expansion frontiers, are not currently captured in global models 13, and is particularly true for models that include mosaic land cover classes. Taken together, our results both show the limits of current data and approaches for predicting cropland globally, and offer some novel and promising ways forward for such work.

**Experimental Procedures**

*Resource Availability*

*Lead Contact*

Further questions about the analysis should be directed to and will be fulfilled by the Lead Contact, Felix Eigenbrod ([f.eigenbrod@soton.ac.uk](mailto:f.eigenbrod@soton.ac.uk)).

*Materials Availability*

This study did not generate new unique materials*.*

*Data and Code Availability*

The code used for all analyses generated during this study, as well as .tif files of frontierness for 2001 and 2015 are freely available on Mendeley Data: <http://dx.doi.org/10.17632/rtgdj2jmk6.1>55

*Methods*

*Cropland data*

We derived global data on cropland in 1992 and 2015 using the ESA’s CCI 2015 dataset, which provides harmonized land cover for all years globally between 1992 and 2015 26. While the ESA CCI land cover is available at the 300 x 300 m resolution, our analyses were conducted at 5’ x 5’ arc minute resolution (percentage of available land that is cropland), as this is the finest common resolution at which the predictor variables for the models of extent and expansion used here are available. To minimize the effects of errors in the ESA CCI 2015 data 50, we excluded pixels classified as mosaic cropland (classes 30 and 40) and only consider those pixels classified as 100% cropland (classes 10,11, 12,20) as cropland in this analysis. We calculated the increase in cropland between 1992 and 2015 at the 300 x 300 m resolution simply by identifying pixels classified as cropland as outlined above in 2015, but not classified as cropland in 1992. We then calculated the percentage cropland cover (for 1992) and change in cropland cover (1992 to 2015) per 5’ x 5’ arcminute pixel (approximately 10 km x 10 km at the equator), using the ESA CCI 150 x 150 m resolution water mask. We then converted % cropland per 5’ x 5’ pixel to a binary variable, as with other recent work 27,56 as the extremely zero-inflated nature of the 1992-2015 expansion data precludes statistical analysis of the raw percentage cropland data. As the threshold used to binarize the cropland variable greatly affects the number of recent cropland expansion pixels globally (see Supplementary Information; Fig S1), we used three different thresholds both for extent and for expansion – > 0.5% cropland, > 10% cropland and > 50% cropland. Pixels classified as having undergone expansion had an increase in cropland corresponding at least to these thresholds between 1992 and 2015.

As all global cropland maps have considerable error associated with them50, we also ran a series of sensitivity analyses to assess whether our findings were robust to how cropland was measured. Firstly, we re-ran the ‘extent’ and ‘expansion’ analyses, using a reclassified ESA CCI map that reclassed the mosaic land cover classes (classes 30 and 40) as cropland. This is because excluding these classes – as is the case in the main analysis as outlined above – could underestimate the total amount (1992) and expansion (1992-2015) of cropland. This could have disproportionately affected smallholder systems, and hence smallholder expansion-driven frontiers. Secondly, we checked the robustness of our results to the choice of end date used for the ‘expansion’ analysis, by re-running all ‘expansion’ analyses using an end date of 2013 rather than 2015. Finally, as land cover products vary enormously in their cropland estimates50, we also re-ran all ‘extent’ and ‘expansion’ models for both the ESA CCI dataset (used in the main analysis), and the MODIS 6 land cover product49 over the period (2001-2015) for which time series are available for both datasets. These datasets represent opposite extremes in estimates of cropland globally, with MODIS generally underestimating cropland, and ESA CCI data prone to overestimating cropland50. Full details of all sensitivity tests are in the Supplementary Experimental Procedures.

*Socio-economic and biophysical explanatory datasets*

We assembled global spatial datasets independent of land cover relevant for cropland extent and expansion which are available at 5’ x 5’ resolution or finer (Table 1). This is because using data modelled using land cover to explain a land cover type (cropland) would lead to circularity in our inference 57. These are: 1) *Bioclimatic suitability*. This is the summed, standardized agroclimatic potential for low-input, rainfed agriculture of the main crops globally - wheat, soy, maize and rice, as well as oil palm 58. The latter is included due to its recent disproportionate role in driving recent agricultural expansion in sub-Saharan Africa 34. Note this excludes edaphic constraints, as these are partially based on land cover. 2) *Steepness* - the percentage of each pixel with a slope of >15% - this is the limit to mechanized agriculture 58. 3*) Access.* Distance to markets 52. 4*) Population density* (in 1990) 59. 5) *GDP*. GDP in 1992 60. We do not consider dynamic variables (e.g. change in population density or GDP) in any models as 1) no relevant gridded global datasets exist prior to 1990; and 2) using dynamic data between 1990 and 2015 to predict expansion between 1990 and 2015 would be circular.

We then built statistical models that linked the above variables to *a priori* theories of cropland expansion (Table 1). Bioclimatic suitability and steepness relates to Ricardian 61 land rent theory, which states that ‘rent’ (underlying value) of land for agriculture is related to its biophysical characteristics relative to other land under use. Distance to markets captures von Thünen’s location theory 62 that the highest land rent is near markets, which makes it most likely to be converted to agriculture. We captured the interactions between the biophysical (Ricardian) and location-based determinants of land rent for agriculture using interaction terms (Equation 1). These simple models do not encompass other key determinants of cropland expansion in land rent theory – high agricultural prices and low alternative employment opportunities 63, as such factors are not available globally in spatially-disaggregated form. We also included human population density, which relates to Boserup’s (1965) theory arguing that expansion of agriculture occurs to satisfy subsistence needs until land becomes scarce, after which intensification takes over. Of course, population expansion can also occur after cropland is established. Unfortunately, high-resolution gridded population data does not exist earlier than 1990, so we were unable to account for this problem of potential endogeneity. Grid cells with high population density should – all other things being equal (modelled via the biophysical constraints, access to markets and interactions between population, access, GDP in 1992 and biophysical constraints; Equation 1) – be dominated by agriculture. By including interactions between economic activity (as measured by GDP) and access, steepness, as well as agroclimatic potential (Equation 1), we incorporated hypotheses concerning potential interactions between subsistence and market demands for agriculture 20. However, these variables do not take into account the technological, institutional variables that also determine demand per unit area – (induced intensification theory), nor quality of governance, inequality in access to land, and other factors that influence the heterogeneity in agent’s ability to materialize potential land rents, all of which are important in determining modern resource frontiers 18,20. Again, this is due to data constraints.

*Explaining extent and expansion with standard theories*

Like-to-like comparisons of the relative importance of different predictors explaining recent agricultural expansion (1992 to 2015) and cropland extent in 1992 requires a sub-sampling approach 15,65. This is because locations of expansion events are comparatively extremely rare to locations where expansions did not occur. Subsampling approaches enable balanced sampling, where an equal number of ‘presences’ (changes) and ‘absences’ are randomly sampled many times from across the study area, and the average coefficients are used 15,65,66. This approach also reduces the spatial autocorrelation 67. We therefore randomly selected 500 ‘presence’ and 500 ‘absence’ squares globally 10,000 times for all three cropland thresholds (0.5%, 10% or 50% or more expansion or extent in cropland) for both extent and expansion models, and then ran a logistic regression models for each subsample.

In each case, we used the same global logistic regression model (Model 1) to enable like-to-like comparisons. This model includes variables that relate to the major existing theories of agricultural expansion 20,63.

*(****Model 1)***

All statistical analyses were limited to areas considered minimally bioclimatically suitable (values > 0) for rain-fed cultivation of at least one of wheat, soy, maize, rice or oil palm by the FAO 58. This excludes deserts, high mountains and the Artic. It also excludes some irrigated croplands in deserts (e.g. Nile valley; Indus valley); this is because spatial data on such irrigated areas is only available using datasets partially based on cropland data 68, which would introduce circularity into the analysis.

In addition, models of expansion were limited to bioclimatically suitable pixels that had sufficient space for agricultural expansion at the relevant threshold. For example, at the 10% threshold, at least 10% of the pixel had to be available for expansion. To limit expansion to realistic pixels 24, the only land within a pixel considered for expansion was areas that were not 1) cropland, urban areas, snow/ice, water or barren land/rock in in 1992; and 2) covered by protected areas (IUCN classes I to IV),69 in 1992.

All variables were centred and standardized globally across suitable pixels, and re-centred for each subsample (as recommended for models with interaction terms) 70. We then calculated the average coefficient, 95% CI and average McFadden pseudo-R2 across all 10,000 models in each instance.

*Positive deviance analysis*

Positive deviance (or ‘bright spot’) analysis identifies locations where the outcome variable exceeds expectations (areas with high positive residuals) derived from a null model. Here, the global ‘null’ model (Null Model) corresponds to all theories of agricultural expansion that can be quantified globally with available data (Table 1); that is, the same predicator variables used for the extent to expansion comparison (Model 1). As existing data do not capture recent frontier expansion well (Experimental Procedures; Q1), positive deviations (areas with more cropland than predicted in 1992) may represent cropland frontier expansion events (frontierness). The Null Models are run for all three cropland thresholds (>0.5%, >10% and >50%) and use all bioclimatically suitable pixels (approximately 1.3 million) because positive deviance analysis cannot be carried via the sub-sampling approach used for Model 1. Our use of positive deviance analysis is somewhat different from other socio-ecological approaches in that our primary interest is not in explaining positive and negative deviations in the 1992 cropland per se, but rather if our putative proxy for expansion frontiers (frontierness; deviations from the null 1992 model) can predict post 1992 cropland expansion events.

***(Null Model****)*

*Does expansion align with positive deviance?*

We quantified the proportional overlap of cropland expansion between 1992 and 2015 at different expansion thresholds (0.5%, 10% and 50% cropland) with areas covered by positive and negative deviations at least 1 SD and 2 SD (respectively) from the global 1992 extent model, following71. Deviation thresholds that had more of their summed total value overlapping with cropland expansion at a given threshold than expected for the area of the cropland have values > 1; deviation thresholds that are under-represented by a given cropland expansion dataset have values below 1. In every case, we used the same cropland threshold (0.5%, 10% and 50%) for both the Null Model and the extent of cropland expansion.

To better understand what the deviations from the 1992 null model (frontierness) represents, we also did post-hoc overlap analyses of frontierness and spatial variables associated with frontier expansion. This approach has previously been used negative deviance (bright spot) analyses in environmental studies30,31. The variables we considered include: i) the human development index (HDI) in 199260; ii) political stability (a country-level indicator from the Global Governance Index [only political stability was chosen (sensu72) as all six Global Governance indices are highly correlated]; iii) the distribution of the largest and smallest quartiles of field size41, and iv) soil constraints on plant growth58. The soil constraint layer comprised binary pixels (5’ x 5’ min resolution); those classified as ‘1’have none of the seven constraints on plant growth in the FAO dataset classified as higher than ‘moderate’; areas classified as ‘0’ have at least one constraint of ‘severe’ or above. None of these variables were suitable for the regression models used in the main analysis, being either too coarse in resolution (political stability), partially derived from land cover (field size, soil constraints) or highly correlated with an existing variable (HDI is highly correlated with GDP), but all link to existing theories of frontier expansion20.

*Does frontierness improve models of expansion?*

We used logistic regression to formally test whether frontierness (deviations from the null 1992 cropland model) represents a useful proxy for expansion frontiers between 1992 and 2015. Our approach is analogous to that of a recent study that uses deviations from a null (biophysical only) model as a new variable for quantifying the impacts of technological advances on banana production globally 73. The use of residuals (or deviations) as a predictors is generally considered poor practice if applied to the same response model 74; however, this is not the case here; cropland in 1992 is independent of expansion of cropland between 1992 and 2015.

To test if frontierness provides additional explanatory power over existing datasets, we compared two different models for each of the three cropland thresholds (0.5%, 10% and 50%). The first –Model 2 – includes the terms (X1) as in Model 1 (the extent to expansion comparison) and the Null Model – as are all globally available independent predictors of expansion – as well as the percentage of cropland in 1992. We include the latter as a predictor, as current extent of cropland is known to be a very good predictor of future expansion, and is therefore incorporated into most land use change models12. As such, Model 2 represents the current state of the art in terms of statistical modelling of global cropland expansion. The third model – Model 3, includes all variables in Model 2, as well as frontierness. We include interaction terms between cropland in 1992 and all other independent predictors as all could potentially moderate the influence of existing cropland on expansion in both Model 2 and Model 3. We also include the interaction term between cropland in 1992 and frontierness in Model 3, as areas with existing cropland in frontier areas could plausibly have higher likelihood of expansion as other areas. We do not include interactions between frontierness and the other independent predictors, as we do not have *a priori* justification for doing so.

The models are:

*(****Model 2****)*

*(****Model 3****)*

All analyses were conducted using R 3.6 75, using a variety of packages.

**Acknowledgements** FE, LJG, RS, and RAH were funded by the ERC Starting Grant “SCALEFORES” (Grant number 680176). SD was funded by “Addressing the Valuation of Energy and Nature Together (ADVENT), Grant/Award Number: NE/M019640/1; NERC, and a University of Southampton studentship. MB acknowledges funding by the Helmholtz Research School for Ecosystem Services under Changing Land Use and Climate (ESCALATE, VH-KO-613). PM was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant agreement No 677140 MIDLAND). XPS acknowledges support by the NASA Harvest Program (80NSSC18M0039). TV was supported by the European Structural and Investments Funds (grant CZ.02.1.01/0.0/0.0/16\_019/0000797: SustES—Adaptation strategies for sustainable ecosystem services and food security under adverse environmental conditions). This work contributes to the Global Land Programme (<https://glp.earth>). The authors also thank two anonymous reviewers whose thoughtful and constructive comments greatly improved this study.

**Author contributions** FE conceived the original idea, ran all analyses and wrote the paper. MB, LG, RAH, PM, RaS, ReS, PV and TV all assisted FE in refining the idea and analyses, and text. SD and XPS contributed data. All authors commented on various version of the final manuscript.

**Declaration of Interests** The authors declare no competing interests.

**References**

1. Ellis, E.C., Kaplan, J.O., Fuller, D.Q., Vavrus, S., Klein Goldewijk, K., and Verburg, P.H. (2013). Used planet: A global history. Proc. Natl. Acad. Sci. *110*, 7978–7985.

2. Haberl, H., Erb, K.H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich, S., Lucht, W., and Fischer-Kowalski, M. (2007). Quantifying and mapping the human appropriation of net primary production in earth’s terrestrial ecosystems. Proc. Natl. Acad. Sci. U. S. A. *104*, 12942–7.

3. Griscom, B.W., Adams, J., Ellis, P.W., Houghton, R.A., Lomax, G., Miteva, D.A., Schlesinger, W.H., Shoch, D., Siikamäki, J. V., Smith, P., et al. (2017). Natural climate solutions. Proc. Natl. Acad. Sci. *114*, 11645–11650.

4. Laurance, W.F., Sayer, J., and Cassman, K.G. (2014). Agricultural expansion and its impacts on tropical nature. Trends Ecol. Evol. *29*, 107–116.

5. Powers, R.P., and Jetz, W. (2019). Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. Nat. Clim. Chang. *9*, 323–329.

6. Zabel, F., Delzeit, R., Schneider, J.M., Seppelt, R., Mauser, W., and Václavík, T. (2019). Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. Nat. Commun. *10*, 1–10.

7. Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. Science *327*, 812–8.

8. Stevenson, J.R., Villoria, N., Byerlee, D., Kelley, T., and Maredia, M. (2013). Green Revolution research saved an estimated 18 to 27 million hectares from being brought into agricultural production. Proc. Natl. Acad. Sci. *110*, 8363–8368.

9. Seppelt, R., Manceur, A.M., Liu, J., Fenichel, E.P., and Klotz, S. (2014). Synchronized peak-rate years of global resources use. Ecol. Soc. *19*, art50.

10. Grassini, P., Eskridge, K.M., and Cassman, K.G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat. Commun. *4*, 1–11.

11. Alexander, P., Prestele, R., Verburg, P.H., Arneth, A., Baranzelli, C., Batista e Silva, F., Brown, C., Butler, A., Calvin, K., Dendoncker, N., et al. (2017). Assessing uncertainties in land cover projections. Glob. Chang. Biol. *23*, 767–781.

12. Prestele, R., Alexander, P., Rounsevell, M.D.A., Arneth, A., Calvin, K., Doelman, J., Eitelberg, D.A., Engström, K., Fujimori, S., Hasegawa, T., et al. (2016). Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison. Glob. Chang. Biol. *22*, 3967–3983.

13. Stehfest, E., van Zeist, W.J., Valin, H., Havlik, P., Popp, A., Kyle, P., Tabeau, A., Mason-D’Croz, D., Hasegawa, T., Bodirsky, B.L., et al. (2019). Key determinants of global land-use projections. Nat. Commun. *10*, 1–10.

14. Delzeit, R., Zabel, F., Meyer, C., and Václavík, T. (2017). Addressing future trade-offs between biodiversity and cropland expansion to improve food security. Reg. Environ. Chang. *17*, 1429–1441.

15. Hatna, E., and Bakker, M.M. (2011). Abandonment and Expansion of Arable Land in Europe. Ecosystems *14*, 720–731.

16. Barbier, E.B. (2012). Scarcity, frontiers and development. Geogr. J. *178*, 110–122.

17. Barbier, E.B. (2020). Long run agricultural land expansion, booms and busts. Land use policy *93*, 103808.

18. le Polain de Waroux, Y., Baumann, M., Gasparri, N.I., Gavier-Pizarro, G., Godar, J., Kuemmerle, T., Müller, R., Vázquez, F., Volante, J.N., and Meyfroidt, P. (2018). Rents, Actors, and the Expansion of Commodity Frontiers in the Gran Chaco. Ann. Am. Assoc. Geogr. *108*, 204–225.

19. de Janvry, A., Fafchamps, M., and Sadoulet, E. (1991). Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained. Econ. J. *101*, 1400–1417.

20. Meyfroidt, P., Roy Chowdhury, R., de Bremond, A., Ellis, E.C., Erb, K.H., Filatova, T., Garrett, R.D., Grove, J.M., Heinimann, A., Kuemmerle, T., et al. (2018). Middle-range theories of land system change. Glob. Environ. Chang. *53*, 52–67.

21. Lambin, E.F., and Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. Proc. Natl. Acad. Sci. U. S. A. *108*, 3465–72.

22. Hertel, T.W., Baldos, U.L.C., and van der Mensbrugghe, D. (2016). Predicting Long-Term Food Demand, Cropland Use, and Prices. Annu. Rev. Resour. Econ. *8*, 417–441.

23. Verburg, P.H., Alexander, P., Evans, T., Magliocca, N.R., Malek, Z., Rounsevell, M. DA, and van Vliet, J. (2019). Beyond land cover change: towards a new generation of land use models. Curr. Opin. Environ. Sustain. *38*, 77–85.

24. Eitelberg, D.A., van Vliet, J., and Verburg, P.H. (2015). A review of global potentially available cropland estimates and their consequences for model-based assessments. Glob. Chang. Biol. *21*, 1236–1248.

25. Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., et al. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science (80-. ). *342*, 850–853.

26. Defourny, P., Kirches, G., Brockmann, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., and Santoro, M. (2016). Land cover CCI product user guide version 2.

27. Curtis, P.G., Slay, C.M., Harris, N.L., Tyukavina, A., and Hansen, M.C. (2018). Classifying drivers of global forest loss. Science (80-. ). *361*, 1108–1111.

28. Hertel, T.W., West, T.A.P., Börner, J., and Villoria, N.B. (2019). A review of global-local-global linkages in economic land-use/cover change models. Environ. Res. Lett. *14*, 053003.

29. Levinson, F.J., Barney, J., Bassett, L., and Schultink, W. (2007). Utilization of positive deviance analysis in evaluating community-based nutrition programs: An application to the Dular program in Bihar, India. Food Nutr. Bull. *28*, 259–265.

30. Cinner, J.E., Huchery, C., MacNeil, M.A., Graham, N.A.J., McClanahan, T.R., Maina, J., Maire, E., Kittinger, J.N., Hicks, C.C., Mora, C., et al. (2016). Bright spots among the world’s coral reefs. Nature *535*, 416–419.

31. Frei, B., Renard, D., Mitchell, M.G.E., Seufert, V., Chaplin-Kramer, R., Rhemtulla, J.M., and Bennett, E.M. (2018). Bright spots in agricultural landscapes: Identifying areas exceeding expectations for multifunctionality and biodiversity. J. Appl. Ecol. *55*, 2731–2743.

32. Alexander, P., Rabin, S., Anthoni, P., Henry, R., Pugh, T.A.M., Rounsevell, M.D.A., and Arneth, A. (2018). Adaptation of global land use and management intensity to changes in climate and atmospheric carbon dioxide. Glob. Chang. Biol. *24*, 2791–2809.

33. Graesser, J., Ramankutty, N., and Coomes, O.T. (2018). Increasing expansion of large-scale crop production onto deforested land in sub-Andean South America. Environ. Res. Lett. *13*.

34. Ordway, E.M., Naylor, R.L., Nkongho, R.N., and Lambin, E.F. (2019). Oil palm expansion and deforestation in Southwest Cameroon associated with proliferation of informal mills. Nat. Commun. *10*, 1–11.

35. Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., and Chookolingo, B. (2018). How much of the world’s food do smallholders produce? Glob. Food Sec. *17*, 64–72.

36. Meyfroidt, P., Schierhorn, F., Prishchepov, A. V., Müller, D., and Kuemmerle, T. (2016). Drivers, constraints and trade-offs associated with recultivating abandoned cropland in Russia, Ukraine and Kazakhstan. Glob. Environ. Chang. *37*, 1–15.

37. Kraemer, R., Prishchepov, A. V., Müller, D., Kuemmerle, T., Radeloff, V.C., Dara, A., Terekhov, A., and Frühauf, M. (2015). Long-term agricultural land-cover change and potential for cropland expansion in the former Virgin Lands area of Kazakhstan. Environ. Res. Lett. *10*, 054012.

38. Dara, A., Baumann, M., Kuemmerle, T., Pflugmacher, D., Rabe, A., Griffiths, P., Hölzel, N., Kamp, J., Freitag, M., and Hostert, P. (2018). Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series. Remote Sens. Environ. *213*, 49–60.

39. Song, X.P., Hansen, M.C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E.F., and Townshend, J.R. (2018). Global land change from 1982 to 2016. Nature *560*, 639–643.

40. Garrett, R.D., Lambin, E.F., and Naylor, R.L. (2013). The new economic geography of land use change: Supply chain configurations and land use in the Brazilian Amazon. Land use policy *34*, 265–275.

41. Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., et al. (2019). Estimating the global distribution of field size using crowdsourcing. Glob. Chang. Biol. *25*, 174–186.

42. Villoria, N.B., and Liu, J. (2018). Using spatially explicit data to improve our understanding of land supply responses: An application to the cropland effects of global sustainable irrigation in the Americas. Land use policy *75*, 411–419.

43. Chen, B., Kennedy, C.M., and Xu, B. (2019). Effective moratoria on land acquisitions reduce tropical deforestation: evidence from Indonesia. Environ. Res. Lett. *14*, 044009.

44. Fritz, S., See, L., Mccallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., et al. (2015). Mapping global cropland and field size. Glob. Chang. Biol. *21*, 1980–1992.

45. Meyfroidt, P. (2018). Trade-offs between environment and livelihoods: Bridging the global land use and food security discussions. Glob. Food Sec. *16*, 9–16.

46. le Polain de Waroux, Y., Garrett, R.D., Heilmayr, R., and Lambin, E.F. (2016). Land-use policies and corporate investments in agriculture in the Gran Chaco and Chiquitano. Proc. Natl. Acad. Sci. *113*, 4021–4026.

47. Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature *553*, 333–336.

48. Fehlenberg, V., Baumann, M., Gasparri, N.I., Piquer-Rodriguez, M., Gavier-Pizarro, G., and Kuemmerle, T. (2017). The role of soybean production as an underlying driver of deforestation in the South American Chaco. Glob. Environ. Chang. *45*, 24–34.

49. Sulla-Menashe, D., and Friedl, M.A. (2018). User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. 1–18.

50. Pérez-Hoyos, A., Rembold, F., Kerdiles, H., and Gallego, J. (2017). Comparison of global land cover datasets for cropland monitoring. Remote Sens. *9*.

51. Boserup, E. (1965). The Conditions of Agricultural Growth: The Economics of Agrarian Societies (Allen & Unwin).

52. Verburg, P.H., Ellis, E.C., and Letourneau, A. (2011). A global assessment of market accessibility and market influence for global environmental change studies. Environ. Res. Lett. *6*, 034019.

53. Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., et al. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. *111*, 3268–3273.

54. Hof, C., Voskamp, A., Biber, M.F., Böhning-Gaese, K., Engelhardt, E.K., Niamir, A., Willis, S.G., and Hickler, T. (2018). Bioenergy cropland expansion may offset positive effects of climate change mitigation for global vertebrate diversity. Proc. Natl. Acad. Sci. *115*, 13294–13299.

55. Eigenbrod, Felix; Beckmann, Michael; Dunnett, Seb; Graham, Laura; Holland, Robert; Meyfroidt, Patrick; Seppelt, Ralf; Song, Xiao-Peng; Spake, Rebecca; Václavík, Tomáš; Verburg, P. (2020). Identifying agricultural frontiers for modelling global cropland expansion - code and data. Mendeley Data. http://dx.doi.org/10.17632/rtgdj2jmk6.1.

56. Oldekop, J.A., Sims, K.R.E., Whittingham, M.J., and Agrawal, A. (2018). An upside to globalization: International outmigration drives reforestation in Nepal. Glob. Environ. Chang. *52*, 66–74.

57. Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., et al. (2013). Challenges and opportunities in mapping land use intensity globally. Curr. Opin. Environ. Sustain. *5*, 484–493.

58. Fischer, G., Nachtergaele, F., Prieler, S., Velthuizen, H.T. van, Verelst, L., and Wiberg, D. (2008). Global Agro-ecological Zones :Model documentation. Food Agric. Organ. United Nations, 179.

59. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid, Future Estimates (2005).

60. Kummu, M., Taka, M., and Guillaume, J.H.A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. Sci. Data *5*, 1–15.

61. Ricardo, D. (1817). The Principles of Political Economy and Taxation (John Murray).

62. von Thunen, J. (1966). The Isolated State (Pergamon).

63. Angelsen, A. (2010). Policies for reduced deforestation and their impact on agricultural production. Proc. Natl. Acad. Sci. *107*, 19639–19644.

64. Boserup, E. (1965). The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure (Allen & Unwin).

65. Levers, C., Schneider, M., Prishchepov, A. V., Estel, S., and Kuemmerle, T. (2018). Spatial variation in determinants of agricultural land abandonment in Europe. Sci. Total Environ. *644*, 95–111.

66. van Asselen, S., and Verburg, P.H. (2012). A Land System representation for global assessments and land-use modeling. Glob. Chang. Biol. *18*, 3125–3148.

67. Sandel, B., and Svenning, J.C. (2013). Human impacts drive a global topographic signature in tree cover. Nat. Commun. *4*, 1–7.

68. FAO-AQUASTAT (2016). Food Agric. Organ. United Nations (FAO). AQUASTAT Database. http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en.

69. Bingham, H.C., Deguignet, M., Lewis, E., Stewart, J., Juffe-Bignoli, D., MacSharry, B., Milam, A., and Kingston, N. (2019). User Manual for the World Database on Protected Areas and world database on other effective area-based conservation measures: 1.6.

70. Dalal, D.K., and Zickar, M.J. (2012). Some Common Myths About Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression. Organ. Res. Methods *15*, 339–362.

71. Eigenbrod, F., Anderson, B.J., Armsworth, P.R., Heinemeyer, A., Jackson, S.F., Parnell, M., Thomas, C.D., and Gaston, K.J. (2009). Ecosystem service benefits of contrasting conservation strategies in a human-dominated region. Proc. R. Soc. B Biol. Sci. *276*, 2903–2911.

72. Václavík, T., Lautenbach, S., Kuemmerle, T., and Seppelt, R. (2013). Mapping global land system archetypes. Glob. Environ. Chang. *23*, 1637–1647.

73. Varma, V., and Bebber, D.P. (2019). Climate change impacts on banana yields around the world. Nat. Clim. Chang. *9*, 752–757.

74. Freckleton, R.P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. J. Anim. Ecol. *71*, 542–545.

75. R Core Team (2019). R: A language and environment for statistical computing.

**Fig 1: Explanatory power of standard theories for extent and expansion***.* Comparison of the explanatory power of global datasets aligned with classical theories of cropland expansion (Table 1) for the global extent of cropland (1992) and recent expansion of cropland (2015) (Model 1; Experimental Procedures**).** Average coefficients for all predictor variables in Model 1 and McFadden pseudo-R2 (± SD) of 10,000 balanced samples (each with 500 cropland and 500 non-cropland pixels) for three different binary thresholds (0.5%, 10% and 50%) of minimum levels of cropland per 5’ x 5’ pixel are shown (Experimental Procedures). Interactions between predictors are shown using “:”; e.g. Population density 1990:GDP.

**Fig 2 –** **Global overlaps of expansion and frontierness.** Theoverlap of recent cropland expansion (1992-2015) and high frontierness (> 1 positive deviations) from global 1992 Null Models of cropland extent at the 0.5% (top panel) and 10% (bottom panel) threshold of minimum levels of cropland per 5’ x 5’ pixel (Experimental Procedures). Results using both thresholds are shown to highlight differences in the distribution of expansion frontiers and how these overlap with the > 1 SD positive deviations. Global results are shown; close-ups from key cropland expansion frontiers (boxes) are shown in Fig 3. Results for the 50% expansion threshold are not shown here as they are too rare to visualize at the global scale. See Fig S2 in the SI for global distributions of extent and expansion of cropland at the >0.5%, >10% and >50% binary expansion thresholds.

**Fig 3 – Overlaps of expansion and frontierness for key expansion frontiers.** Theoverlap of recent cropland expansion (1992-2015) and large (> 1 SD) positive deviations from global 1992 Null Models of cropland extent at the 0.5% (top panel) and 10% (bottom panel) threshold (Experimental Procedures) for key cropland expansion frontiers. The letters correspond to the locations of the regions on the global map (Fig 2).

**Table 1: Independent global predictor variables and their theoretical justification for cropland expansion.** These predictor variables were used to explain cropland extent in 1992 and expansion between 1992-2015 (Model 1), and construct the Null Model of current cropland extent. The \* represents interaction terms between variables. Full details of the justification for these variables are in the Experimental Procedures.

|  |  |
| --- | --- |
| **Variable** | **Theoretical justification** |
| Bioclimatic suitability58 | Ricardian land rent theory 61 |
| Steepness58 | Ricardian land rent theory 61 |
| Access (distance to markets)52 | von Thünen’s location theory 62 |
| Bioclimatic suitability \* Access OR Steepness \* Access | Interactions between land rent theory and location theory (i.e. steep land is more valuable near markets) |
| Population Density (1990)59 | Induced intensification 64 |
| GDP (1992)60 | Market demands for agriculture 20 |
| Population Density \* bioclimatic suitability OR Steepness OR Access OR GDP | Interactions between induced intensification, land rent theory and between induced intensification and location theory or market vs subsistence demands for agriculture |
| GDP \* Access OR Steepness OR Biophysical constraints | Interactions between subsistence and market demands on agriculture 20 |

**Table 2: Proportional overlap of expansion of cropland between 1992 and 2015** **and deviations from the global ‘null’ model of cropland extent in 1992 (Experimental Procedures; Null Model).** The predictor variables in the Null Model are the same as in Model 1 (Table 1), but the Null Model uses all data globally; Model 1 uses balanced samples of 500 presences and absences (Experimental Procedures). ‘Positive Deviance’ and ‘Negative Deviance’ refer to 2 or 1 or more positive (or negative) standard deviations (2SD and 1SD) from the Null Model. Ratios > 1 indicate over-representation; ratios < 1 indicate under-representation (more or less overlap of cropland and a given deviation threshold than would be expected if both are equally common across the land area they cover 71).

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Expansion of cropland 1992-2015** | | |
|  | **0.5% crop** | **10% crop** | **50% crop** |
| Positive Deviance 2SD | 1.35 | 3.05 | 0.03 |
| Positive Deviance 1SD | 1.53 | 2.58 | 0.03 |
| Negative Deviance 2SD | 0.41 | 0.39 | 2.41 |
| Negative Deviance 1SD | 0.34 | 0.43 | 3.40 |

**Table 3: Average predictive power (McFadden pseudo-R2) of models with and without frontierness for recent (1992 to 2015) expansion of cropland.** Model 2 includes all independent predictors of expansion aligned with classical theories of expansion (that is all predictors in Model 1 and the Null Model; outlined in Table 1), as well as the percentage of cropland in 1992 as a predictor, and its interaction with bioclimatic suitability, access, population density in 1990 and GDP in 1992. The inclusion of existing cropland is a standard practice in global models predicting land use change. Model 3 includes all terms in Model 2, as well as frontierness, and the interaction of frontierness and cropland in 1992 (Experimental Procedures). Results are averages of 1000 balanced subsamples of expansion for each model (Experimental Procedures). Full model results are given in the SI for Model 2 (Table S3) and Model 3 (Table S4).

|  |  |  |
| --- | --- | --- |
| Expansion threshold | Existing predictors (Model 2) | Frontierness + existing predictors  (Model 3) |
| 0.5% cropland | 0.12 | 0.20 |
| 10% cropland | 0.23 | 0.31 |
| 50% cropland | 0.41 | 0.41 |

Figures

A picture containing diagram

Description automatically generated

Figure 1

Diagram

Description automatically generated

Figure 2

Diagram, engineering drawing

Description automatically generated

Figure 3

**Supplementary Information**

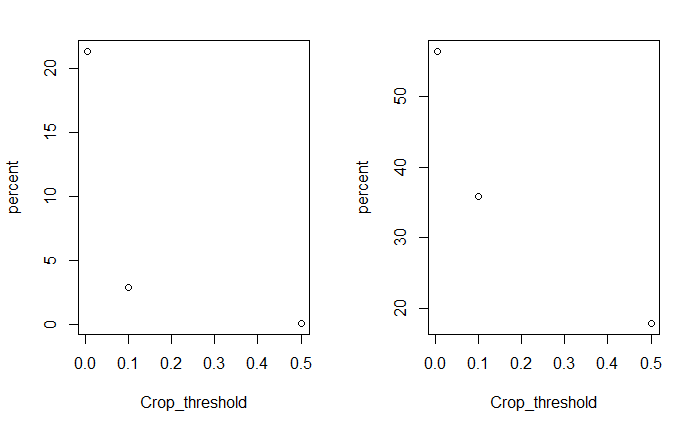
-Supplementary Figures

-Supplementary Experimental Procedures

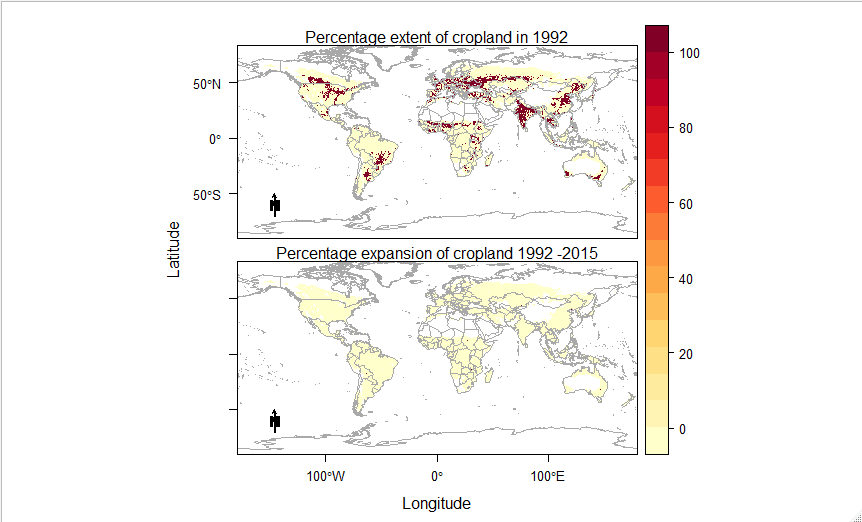
-Supplementary Tables

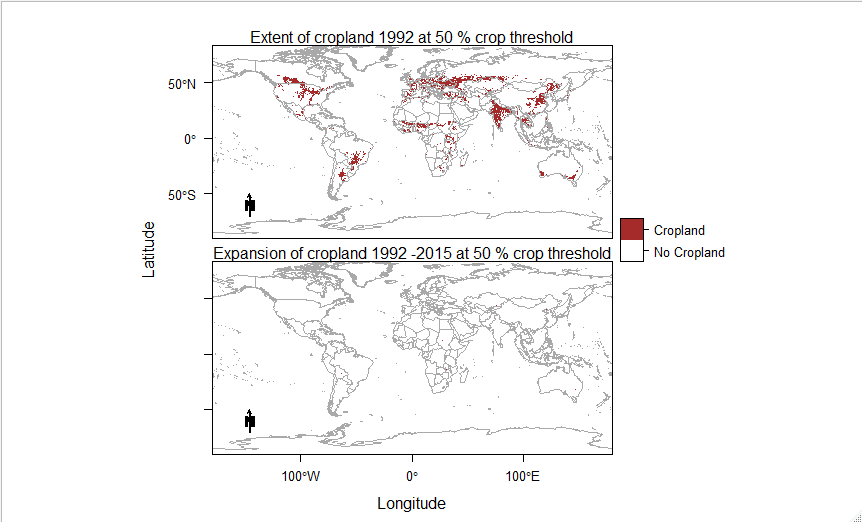
-Supplementary References

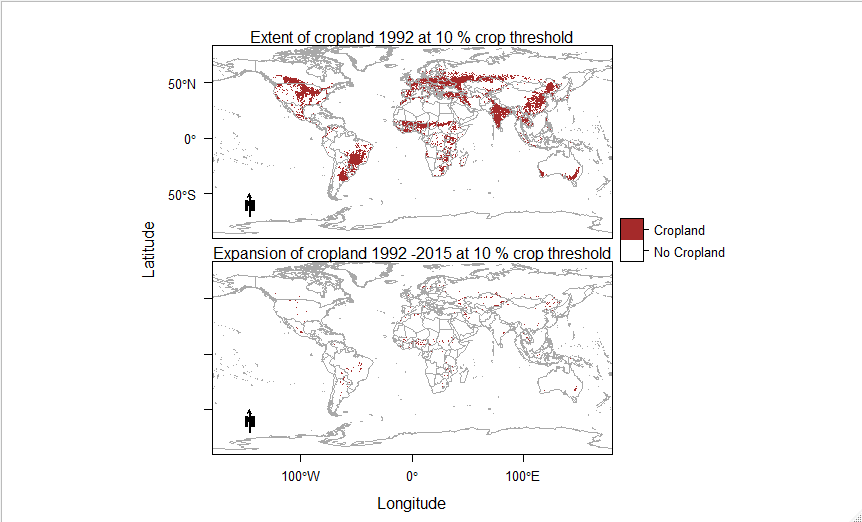
**Supplementary Figures**

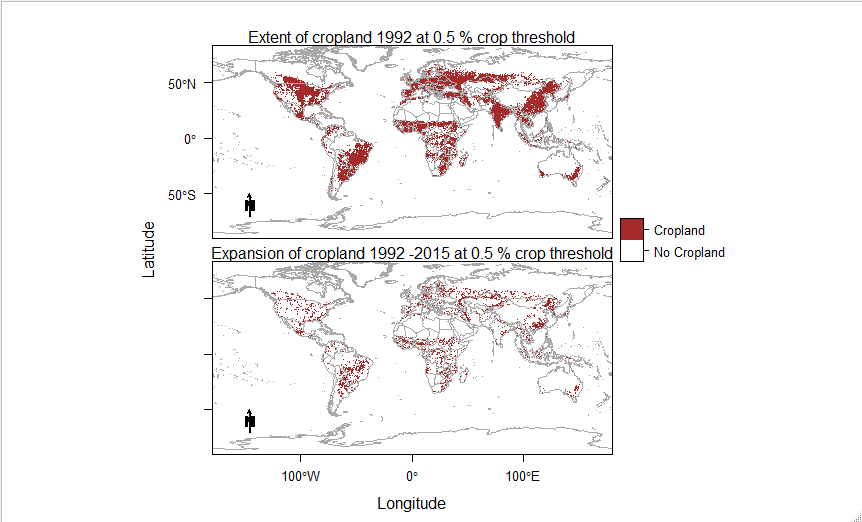


**Fig S1 Relationship between the percentage of bioclimatically suitable land area for cropland and the threshold of the percentage of cropland per 5’ x 5’ arc minute grid cell (Crop\_threshold) for recent expansion (1992 to 2015; left) and cropland extent in 1992; right).** The value near 0 is for a threshold of 0.005 (0.5%) cropland per grid cell. In total, 0.45% of expansion at the 0.5% threshold is also above the 50% cropland threshold; 13.6% of expansion at the 0.5% threshold is also above the 10% cropland threshold. Cropland is based on ESA CCI data and excludes mosaic land cove classes (Methods).

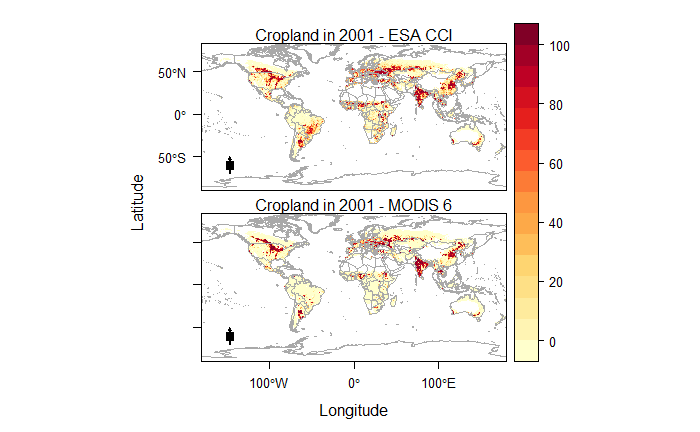




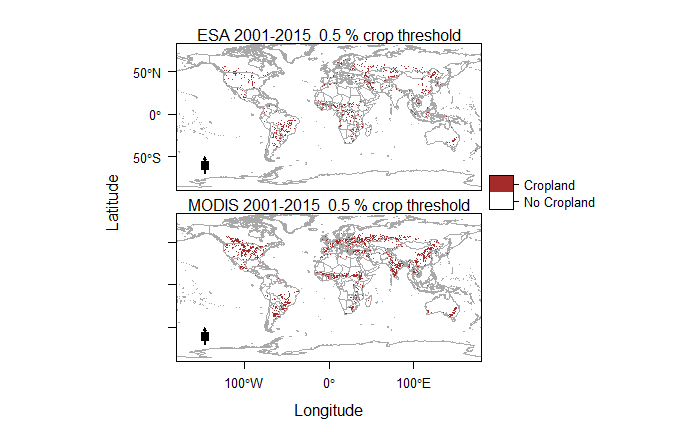


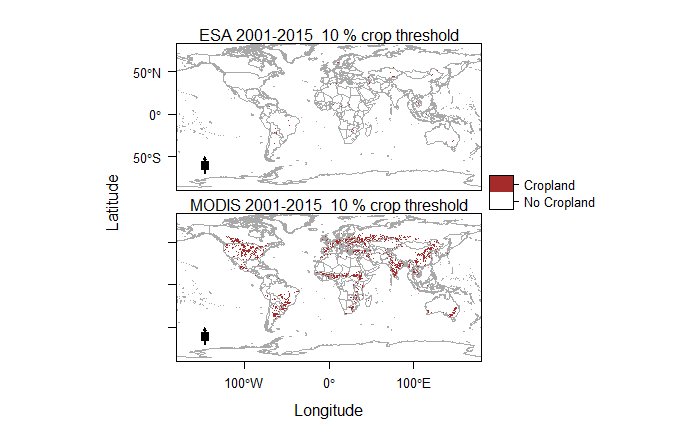


**Fig. S2 Distribution of the global extent (in 1992) and recent expansion (1992-2015) of cropland in terms of percentage per 5’ x 5’ pixel, and at different binary thresholds of cropland per pixel (>0.5%; >10% and >50%).** In all cases, only cropland within the bioclimatic envelope (Methods) for rain-fed agriculture is shown, as this was data used in the analysis. Cropland is based on ESA CCI data and excludes mosaic land cove classes (Methods). Note that expansion events at the 50% threshold are extremely rare (~1300 pixels globally), making it difficult to distinguish them in a global map.



**Fig S3 – Distributions of percentage cropland per 5’ x 5’ pixel in 2001 as measured by the ESA CCI (top) and MODIS 6 (bottom) data.** In both cases, mosaic land cover classes are excluded in the calculations of the distribution of cropland. See Supplementary Methods for details on dataset creation.





**Fig S4: Distribution of expansion (2001-2015) of cropland as measured using the ESA CCI and MODIS 6 land cover products at different binary thresholds of cropland per 5’ x 5’ pixel (>0.5% and >10%).** In all cases, only cropland within the bioclimatic envelope (Methods) for rain-fed agriculture is shown, as this was data used in the analysis. In both cases, mosaic land cover classes are excluded in the calculations of the distribution of cropland. See Supplementary Methods for detailed methods.

**Supplementary Experimental Procedures**

Post-hoc proportional overlap of expansion and other socio-economic variables.

We undertook post-hoc analyses of the proportional overlaps of recent cropland expansion and a number of additional socio-economic variables. This enabled us to better understand what is driving this expansion (Table S6). These were: 1) change in HDI and 2) change in GDP between 1992 and 2015; 1; 3) change in human population between 1990 and 20152, 4) the largest and 5) smallest quartiles of field size)3; 6) percentage irrigated area (4; locations where deforestation between 2003 and 2015 was classified as primarily caused by 7) commercial agriculture and 8) subsistence agriculture 5; and distributions of 9) oil palm and 10) soy6. These analyses were carried out post-hoc, and only via overlap analyses, as none of these datasets are suitable for predictive modelling of cropland expansion. This because socio-economic change data (so change in HDI, GDP and populations) is only relevant for explaining changes that have already occurred, not prediction into the future, while the field size, irrigated area, deforestation, and oil palm and soybean datasets all include cropland data within them, and hence would introduce circularity in statistical testing. Moreover, the cropland data used in the above datasets comes from multiple years between 1992 and 2015. As such, the results of these overlaps should be viewed with considerable caution. However, comparing the overlaps between the different threshold of cropland expansion (0.5%, 10% and 50%) does provide some additional insights of whether expansion at different thresholds is driven mainly by corporate actors (associated with very large fields dominated by cash crops such as soy and oil palm) or smallholders7,8. Variables that had more of their summed total value overlapping with cropland expansion at a given threshold than expected for the area of the cropland have values > 1; Variables that are under-represented by a given cropland expansion dataset have values below 1. This is the same method as used for cropland expansion and deviations from the Null Model (Methods). Results are shown in Table S8.

Post-hoc proportional overlap of forest to short vegetation transitions between 1982 and 1992 and 1992 and 2015 and areas of high positive deviance from the null model

To better understand what the frontierness proxy (deviations from the global Null Model of cropland extent in 1992; Methods) is measuring, we conducted additional post-hoc analyses using a recent long-term time-series of vegetation changes (1982-2016) measured using AVHRR data9. Specifically, we looked at the proportional overlap of transitions from forests to short vegetation between 1982 and 1992 and areas of high deviation (+1 and +2 SD, respectively) using the proportional overlap method outlined above. Our hypothesis is that if these deviations represent recent expansion frontiers, these areas of deforestation would be disproportionately overrepresented by these areas of high deviation. However, the AVHRR data is not directly comparable to the ESA CCI dataset we use for recent land use change. The AVHRR-based forest-to-short vegetation transition represents deforestation for both cropland and pasture expansion9. This can lead to major discrepancies between the approaches. For example, post-2000, new croplands in tropical South America were mostly converted from pasture 10. Indeed, the global Spearman correlation between non-zero shifts from forest to short vegetation and the ESA CCI derived percentage increase in cropland at the 5’ resolution is only 0.18. We therefore also quantified the proportional overlap of the AVHRR data between 1992 and 2015, thereby enabling us to make a like-to-like comparison of how well areas of high deviation from the 1992 Null model overlap with deforestation frontiers before and after 1992.

As breaking the 1982 to 2016 time series used in Song et al into two separate time slices reduces the sample size for per-pixel trend analyses, we ran the 1982-1992 and 1992-2015 analyses at 0.08°, rather than 0.05° in the original analysis. Specifically, we first resampled the annual tree cover (TC) and short vegetation (SV) layers from 0.05° to 0.04° using bilinear resampling and aggregated them from 0.04° to 0.08° – taking the mean value of each 2 x 2 window. We tested non-parametric trends with the aggregated data, separately for TC and SV, following 9. Then, a TC-to-SV transition was identified such that TC showed a statistically significant decreasing trend (p<0.05) and SV showed a statistically significant increasing trend (p<0.05). We computed the Theil-Sen slopes for TC change and SV change, and took the smaller of the two as the annualized TC-to-SV change. We then converted it to total percentage change over the length of the specific time period. This was carried out separately for 1982-1992 and 1992-2015. Results are shown in Table S9.

Sensitivity testing of results to the assumptions of the ESA CCI crop map

In the main analysis, we excluding mosaic land cover classes. This could have disproportionately affected smallholder systems, and hence smallholder expansion driven frontiers. We therefore re-ran our analyses to include cropland represented by mosaic land cover classes, as well as the 100% cropland (10, 11, and 12) land cover classes. The two mosaic land cover classes in the ESA CCI dataset are 30 (>50% cropland) and 40 (<50% cropland). In absence of additional data, we assumed class 30 was 75% cropland, and class 40 25% cropland. Results are shown in Table S5 (‘extent’) and Table S6 (‘expansion’).

As land cover products vary enormously in their estimates of cropland11, we also re-ran all ‘extent’ and ‘expansion’ models for both the ESA CCI dataset (used in the main analysis), and the MODIS 6 land cover product12 over the period (2001-2015) for which time series are available for both datasets. We only considered ‘pure’ cropland classes (100% cropland; classes, 10, 11, and 20 for ESA CCI; class 12 for MODIS 6 to align with the main analysis (distributions shown in Fig S3). We also followed the same method for calculating cropland change as for the main analysis (distributions shown in Fig S4). Specifically, we calculated the increase in cropland between 2001 and 2015 at the 300 x 300 m resolution (for the ESA CCI data) or 500 x 500 m resolution (MODIS 6 data) by identifying pixels classified as cropland as outlined above in 2015, but not classified as cropland in 2001. We then calculated the percentage cropland cover (for 2001) and change in cropland cover (2001 to 2015) per 5’ x 5’ arcminute pixel (approximately 10 km x 10 km at the equator), using the ESA CCI 150 x 150 m resolution water mask (resampled for the MODIS data). We then calculated the land available for crops in 2001 using the same methods as before (though run at 500 x 500 m resolution for the MODIS data), but used world protected area database from 2001 instead of 199213 to exclude protected areas within these calculations. We then re-ran all ‘extent’ (Model 1) and ‘expansion’ (Model 2 and Model 3) analyses, using the same equations by updated key predictors to align as closely as possible to 2001 (as opposed to 1992 in the base analysis). Specifically, instead of using population density in 19902, we used population density in 200014; GDP in 2001 instead of 19921. Frontierness in the ‘expansion’ analysis was always based on the deviations from the corresponding ‘extent’ analysis (e.g. frontierness in the ‘expansion’ analysis using MODIS data for 2001-2015 was based on the ‘extent’ analysis based on MODIS data for 2001).

Finally, we checked to see if our results were robust to potential noise in the specific endpoint we used (2015). Specifically, we re-ran our main ‘expansion’ analysis for the years 1992-2013, rather than 1992-2015 (Table S5).

Creation of geospatial files of frontierness for 1992 and 2015

We provide global 5’ x 5’ resolution geospatial files (.tif) of frontierness for 1992 and 2015, based on the main analysis (ESA CCI data; mosaic land cover classes excluded). Values for both files are standardized between 0 and 1, exclude areas that are not bioclimatically suitable (Methods), and are in geographic projection. The 1992 file is the standardized deviations from the global Null Model, as described in the Methods. Note Population Density is from 1990:

The 2015 file is identical in structure, but with the 2015 cropland data as the response, and updated datasets where these are available (GDP for 2015 and Population Density for 2015). The 2015 GDP data comes from 1 (the same as the 1992 GDP data), while the Population Density data for 2015 comes from SEDAC v4 14 (as opposed to SEDAC v3 for the 1990 data2).

**Supplementary Tables**

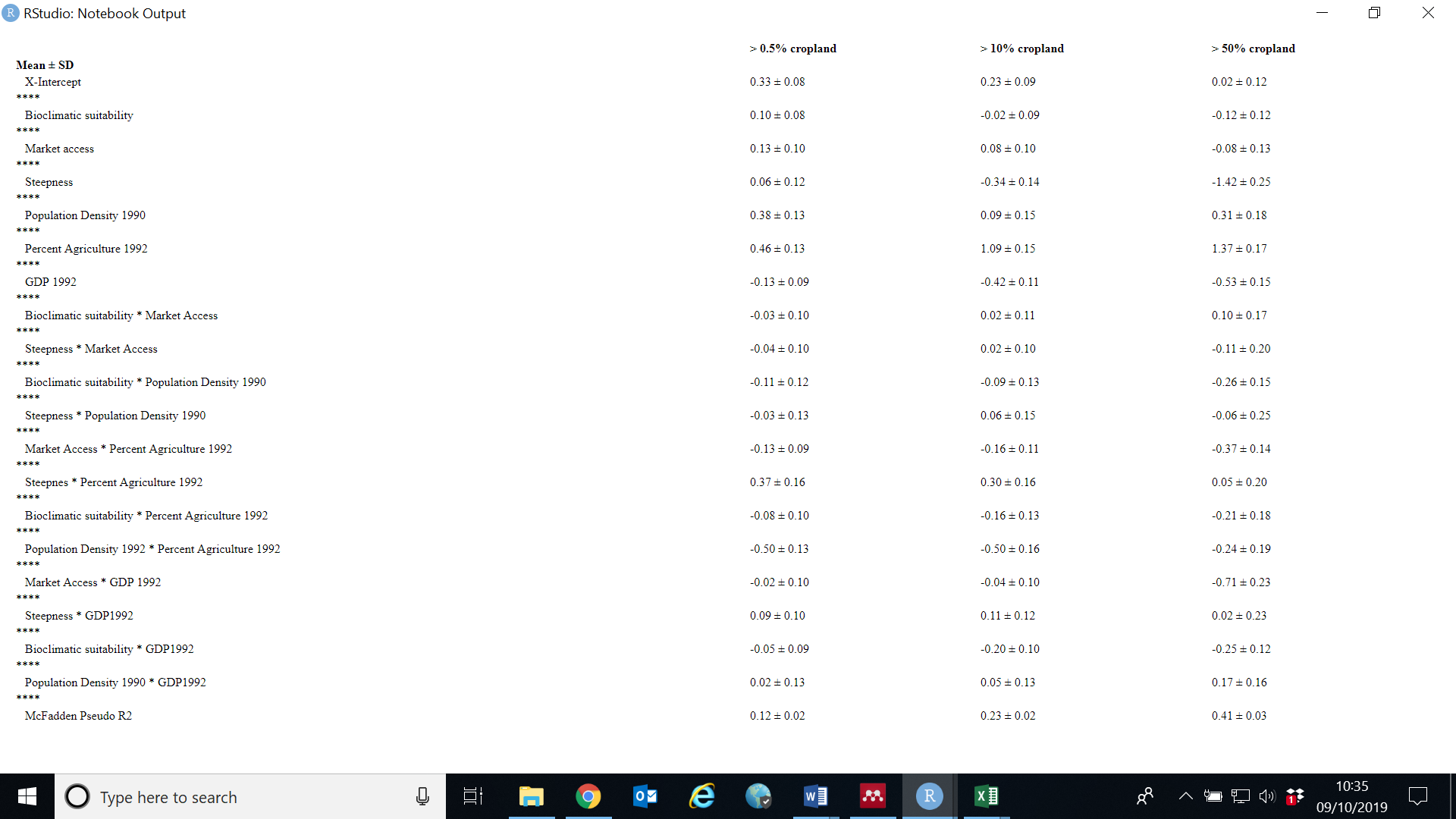
**Table S1 – Model coefficients and McFadden pseudo-R2  of the Null Model (Methods) of global cropland extent in 1992.** The deviations from this model form the basis of the frontierness proxy.

|  |  |  |  |
| --- | --- | --- | --- |
| **Mean ± SE** | **> 0.5% cropland in 1992** | **> 10% cropland in 1992** | **> 50% cropland in 1992** |
| Intercept | 0.495±0.003 | 0.003±-0.894 | -0.894±0.003 |
| Bioclimatic suitability | 0.06±0.003 | 0.003±0.112 | 0.112±0.003 |
| Market access | 0.622±0.005 | 0.005±0.375 | 0.375±0.003 |
| Steepness | -0.355±0.003 | 0.003±-0.733 | -0.733±0.004 |
| Population Density 1990 | 1.597±0.004 | 0.004±1.571 | 1.571±0.005 |
| GDP 1992 | 0.011±0.003 | 0.003±0.14 | 0.14±0.003 |
| Bioclimatic suitability \* Market Access | -0.171±0.004 | 0.004±-0.124 | -0.124±0.003 |
| Steepness \* Market Access | -0.179±0.004 | 0.004±-0.149 | -0.149±0.003 |
| Bioclimatic suitability \* Population Density 1990 | -0.02±0.004 | 0.004±-0.137 | -0.137±0.004 |
| Steepness \* Population Density 1990 | -0.069±0.004 | 0.004±0.044 | 0.044±0.005 |
| Market Access \* GDP 1992 | -0.239±0.004 | 0.004±-0.174 | -0.174±0.003 |
| Steepness \* GDP1992 | -0.053±0.003 | 0.003±-0.028 | -0.028±0.004 |
| Bioclimatic suitability \* GDP1992 | 0.138±0.003 | 0.003±0.107 | 0.107±0.003 |
| Population Density 1990 \* GDP 1992 | -0.266±0.004 | 0.004±-0.423 | -0.423±0.004 |
| McFadden Pseudo-R2 | 0.307±0.307 | 0.307±0.288 | 0.288±0.288 |

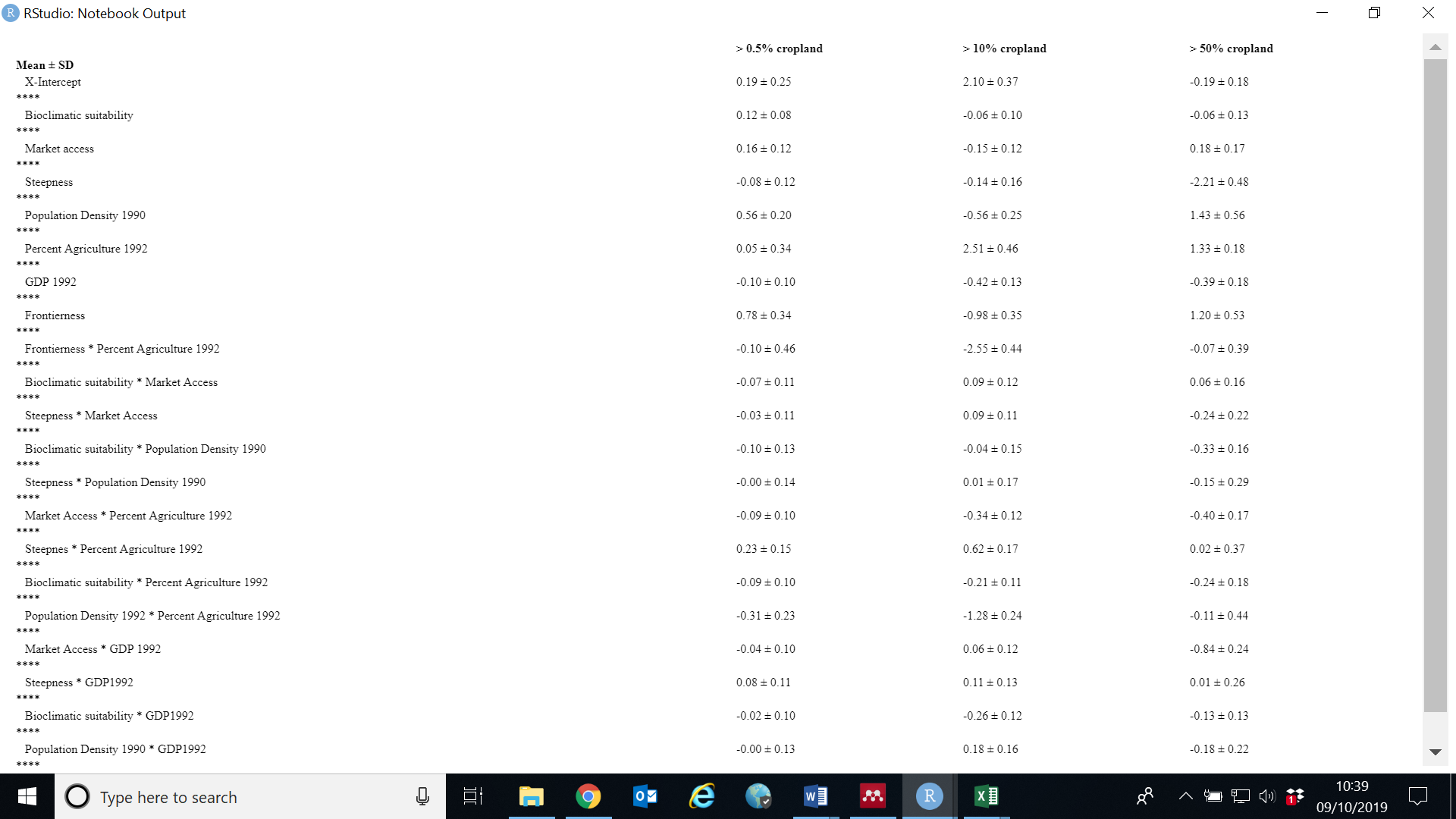
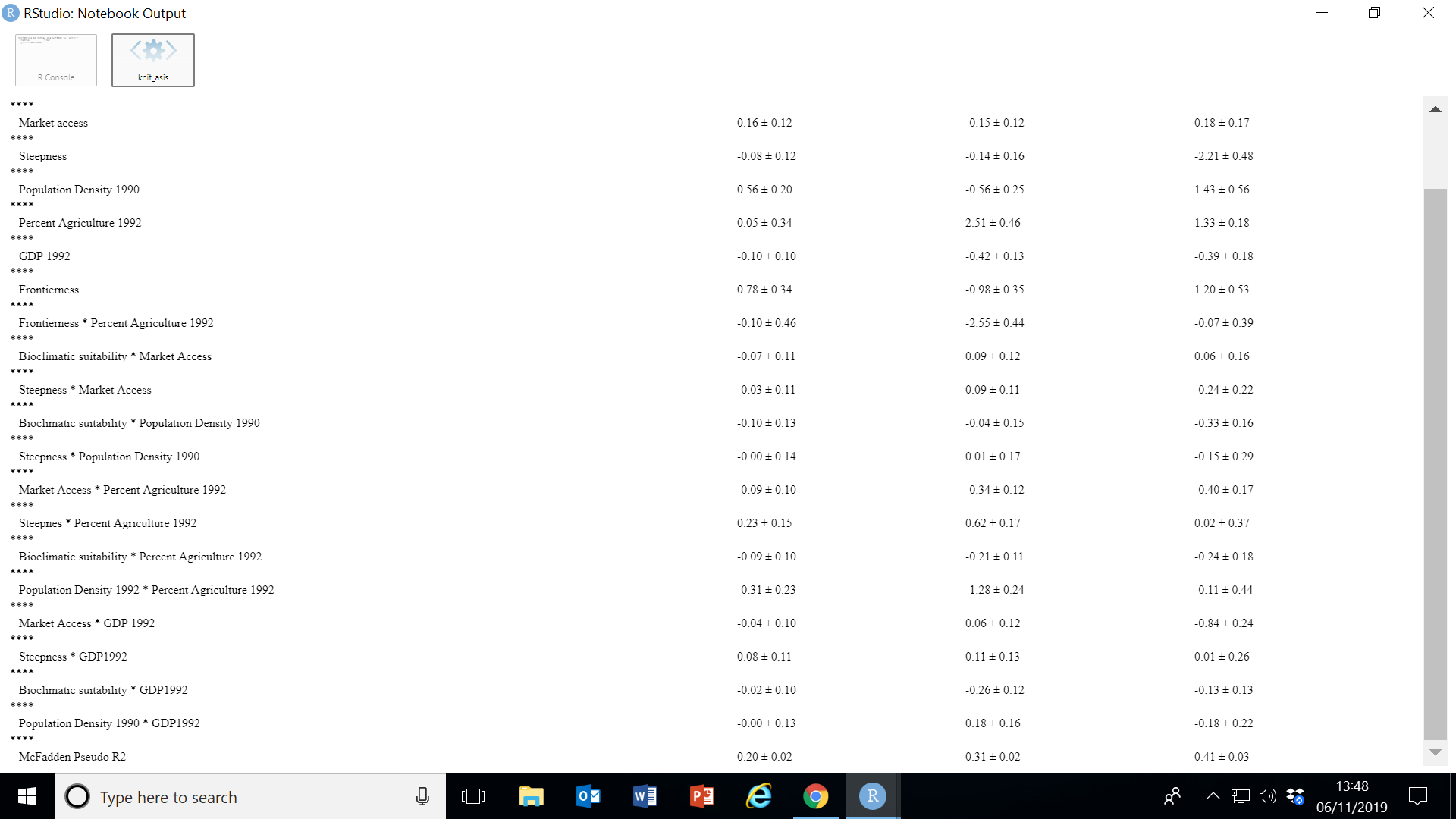
**Table S2 –** Proportional overlap of expansion of variables potentially associated with frontier conditions, and deviations from the global ‘null’ model of cropland extent in 1992 (Experimental Procedures; Null Model). The predictor variables in the Null Model are the same as in Model 1 (Table 1), but the Null Model uses all data globally; Model 1 uses balanced samples of 500 presences and absences (Experimental Procedures). ‘Positive Deviance’ and ‘Negative Deviance’ refer to 2 or 1 or more positive (or negative) standard deviations (2SD and 1SD) from the Null Model. Ratios > 1 indicate over-representation; ratios < 1 indicate under-representation (more or less overlap of given variable and a given deviation threshold than would be expected if both are equally common across the land area they cover 15.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Threshold | Variable | Positive Deviance 2 SD | Positive Deviance 1 SD | Negative Deviance 2 SD | Negative Deviance 1 SD |
| 0.5% | Largest quartile of field sizes | **1.82** | **2.32** | 0.38 | 0.38 |
| 0.5% | Smallest quartile of field sizes | 0.05 | 0.40 | **1.70** | 0.83 |
| 0.5% | HDI in 1992 | **1.22** | **1.05** | **1.03** | 0.88 |
| 0.5% | Political stability | **1.18** | **1.05** | 0.98 | 0.91 |
| 0.5% | Soil constraints | **1.25** | **1.10** | 0.17 | 0.07 |
| 10% | Largest quartile of field sizes | **3.16** | **2.56** | 0.52 | 0.62 |
| 10% | Smallest quartile of field sizes | 0.26 | 0.99 | **2.14** | **1.55** |
| 10% | HDI in 1992 | **1.16** | **1.04** | 0.94 | 0.95 |
| 10% | Political stability | **1.14** | **1.02** | 0.94 | 0.96 |
| 10% | Soil constraints | **1.12** | **1.26** | **1.05** | 0.97 |
| 50% | Largest quartile of field sizes | **3.53** | **2.76** | 0.58 | 0.71 |
| 50% | Smallest quartile of field sizes | 0.73 | **1.15** | **2.83** | **2.03** |
| 50% | HDI in 1992 | **1.15** | **1.08** | 0.87 | 0.98 |
| 50% | Political stability | **1.11** | **1.03** | 0.89 | 0.98 |
| 50% | Soil constraints | **1.36** | **1.47** | **1.17** | **1.06** |

**Table S3 – Average model coefficients and McFadden pseudo-R2  (± SD) of 10,000 balanced samples of expansion (1992-2015) of Model 2 (‘existing predictors’); Methods.**



**Table S4 – Average model coefficients and McFadden pseudo-R2  (± SD) of 10,000 balanced samples of expansion (1992-2015) of Model 3 (‘existing predictors + frontierness’); Methods.**

**Table S5**– Comparison of the overall model explanatory power (pseudo R2) of the main cropland classification (‘main’) for the ‘extent’ models with cropland that includes mosaic land cover classes (‘mosaic’). See Supplementary Methods for Details.

|  |  |  |  |
| --- | --- | --- | --- |
| Model type | Cropland Threshold | Main Analysis | Mosaic |
| Extent | 0.005 | 0.32 | 0.34 |
| Extent | 0.1 | 0.30 | 0.32 |
| Extent | 0.5 | 0.33 | 0.33 |

**Table S6** – Comparisons of the overall model predictive power (pseudo R2) main cropland classification (‘main’) for the ‘expansion’ models with cropland that includes mosaic land cover classes (‘mosaic’), and cropland expansion up to 2013 (‘2013’), rather than up to 2015. See Supplementary Methods for Details.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Model 2 (Cropland + null) | | | Model 3 (Cropland + null + frontierness) | | |
| Expansion | Cropland Threshold | Main Analysis | Mosaic | 2013 | Main Analysis | Mosaic | 2013 |
| Expansion | 0.005 | 0.12 | 0.11 | 0.12 | 0.20 | 0.20 | 0.20 |
| Expansion | 0.1 | 0.23 | 0.16 | 0.23 | 0.31 | 0.32 | 0.32 |
| Expansion | 0.5 | 0.41 | 0.30 | 0.41 | 0.41 | 0.41 | 0.41 |

**Table S7**– Comparison of the overall model explanatory power (pseudo R2) for ‘extent’ models for 2001, and predictive power (pseudo R2) of the ‘expansion’ models (between 2001-2015) based on cropland as quantified by the ESA CCI data (‘ESA CCI’) and MODIS 6 data (‘MODIS’). See Supplementary Methods for Details.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Extent Model (Model 1) | | Model 2 (Cropland + null) | | Model 3 (Cropland + null + frontierness) | |
| Cropland Threshold | ESA CCI | MODIS | ESA CCI | MODIS | ESA CCI | MODIS |
| 0.005 | 0.29 | 0.29 | 0.09 | 0.35 | 0.16 | 0.53 |
| 0.1 | 0.28 | 0.29 | 0.23 | 0.42 | 0.29 | 0.57 |

**Table S8** – Proportional overlap of expansion of cropland between 1992 and 2015, and socio-economic variables of interest. Ratios > 1 indicate over-representation; ratios < 1 indicate under-representation (more or less overlap of cropland and a given variable than would be expected if both are equally common across the land area they cover)15. Details on the variables are in the Suppplementary Methods.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Expansion of cropland 1992-2015 | | |
|  | 0.5% crop | 10% crop | 50% crop |
| CurtisCommercial Agriculture | 1.87 | 2.62 | 3.09 |
| Curtis\_Subsistence Agriculture | 1.18 | 1.27 | 1.08 |
| Largest quartile of field sizes | 1.33 | 1.72 | 2.29 |
| Smallest quartile of field sizes | 1.35 | 1.49 | 1.28 |
| GDP Changes 1992\_2015 | 0.99 | 1.01 | 0.87 |
| HDI Changes 1992\_2015 | 0.99 | 1.04 | 1.01 |
| Change in population density 1990\_2015 | 0.97 | 0.97 | 0.94 |
| Irrigated areas | 1.13 | 1.36 | 0.85 |
| Oil palm cultivation | 1.79 | 1.86 | 1.32 |
| Soy cultivation | 1.29 | 2.18 | 3.32 |

**Table S9: Proportional overlap of shifts from forest to short vegetation (1982-1992 and 1992-2015) and large positive deviations from the global ‘null’ model of cropland extent in 1992 (Methods; Null Model).** ‘Positive Deviances’ of 2 or more (or 1 or more) standard deviations are shown. Ratios > 1 indicate over-representation; ratios < 1 indicate under-representation (more or less overlap of cropland and a given deviation threshold than would be expected if both are equally common across the land area they cover)15. As the positive deviance analysis was conducted using separate Null Models for different percent cropland binary cut-offs (>0.5%, >10%; >50%), separate overlaps were conducted for each thresholds. The shifts from forest to short vegetation are measured using AVHRR data following 9; see Supplementary Methods for details.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Forest to short veg 1982-1992** | | | **Forest to short veg 1992-2015** | | |
|  | 0.5% crop | 10% crop | 50% crop | 0.5% crop | 10% crop | 50% crop |
| Positive Deviance 2 SD | 3.84 | 3.12 | 1.84 | 3.48 | 1.29 | 1.11 |
| Positive Deviance 1 SD | 1.79 | 1.43 | 0.72 | 1.51 | 0.85 | 0.26 |

**Supplementary References**

1. Kummu, M., Taka, M., and Guillaume, J.H.A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. Sci. Data *5*, 1–15.

2. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid, Future Estimates (2005).

3. Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., et al. (2019). Estimating the global distribution of field size using crowdsourcing. Glob. Chang. Biol. *25*, 174–186.

4. FAO-AQUASTAT (2016). Food Agric. Organ. United Nations (FAO). AQUASTAT Database. http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en.

5. Curtis, P.G., Slay, C.M., Harris, N.L., Tyukavina, A., and Hansen, M.C. (2018). Classifying drivers of global forest loss. Science (80-. ). *361*, 1108–1111.

6. You, L., Wood, S., Wood-Sichra, U., and Wu, W. (2014). Generating global crop distribution maps: From census to grid. Agric. Syst. *127*, 53–60.

7. le Polain de Waroux, Y., Garrett, R.D., Graesser, J., Nolte, C., White, C., and Lambin, E.F. (2019). The Restructuring of South American Soy and Beef Production and Trade Under Changing Environmental Regulations. World Dev. *121*, 188–202.

8. le Polain de Waroux, Y., Baumann, M., Gasparri, N.I., Gavier-Pizarro, G., Godar, J., Kuemmerle, T., Müller, R., Vázquez, F., Volante, J.N., and Meyfroidt, P. (2018). Rents, Actors, and the Expansion of Commodity Frontiers in the Gran Chaco. Ann. Am. Assoc. Geogr. *108*, 204–225.

9. Song, X.P., Hansen, M.C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E.F., and Townshend, J.R. (2018). Global land change from 1982 to 2016. Nature *560*, 639–643.

10. Zalles, V., Hansen, M.C., Potapov, P. V., Stehman, S. V., Tyukavina, A., Pickens, A., Song, X.P., Adusei, B., Okpa, C., Aguilar, R., et al. (2019). Near doubling of Brazil’s intensive row crop area since 2000. Proc. Natl. Acad. Sci. U. S. A. *116*, 428–435.

11. Pérez-Hoyos, A., Rembold, F., Kerdiles, H., and Gallego, J. (2017). Comparison of global land cover datasets for cropland monitoring. Remote Sens. *9*.

12. Sulla-Menashe, D., and Friedl, M.A. (2018). User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. 1–18.

13. Bingham, H.C., Deguignet, M., Lewis, E., Stewart, J., Juffe-Bignoli, D., MacSharry, B., Milam, A., and Kingston, N. (2019). User Manual for the World Database on Protected Areas and world database on other effective area-based conservation measures: 1.6.

14. Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O., and Baptista, S.R. (2015). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. Pap. Appl. Geogr. *1*, 226–234.

15. Eigenbrod, F., Anderson, B.J., Armsworth, P.R., Heinemeyer, A., Jackson, S.F., Parnell, M., Thomas, C.D., and Gaston, K.J. (2009). Ecosystem service benefits of contrasting conservation strategies in a human-dominated region. Proc. R. Soc. B Biol. Sci. *276*, 2903–2911.