**Appendix 2.** Supplementary information on the methodological approach

**Land cover classification methodology**

We performed a supervised classification of Landsat images for the years 2000 and 2020. We used the Google Earth Engine platform (Gorelick et al., 2017) because is a cloud online platform that allows processing massive remote sensing data effectively.

We used LANDSAT/LT05/C01/T1\_SR image collection for the year 2000 and LANDSAT/LC08/C01/T1\_SR for the year 2020. Therefore, our source images consist of surface reflectance satellite images. We applied a cloud mask algorithm for image collections using the median value of data per pixel using images from January to December. We filter images from 2000 to 2002 year to compose the image of 2000 and from 2016 to 2020 to compose the image of 2020.

Since we are working with a large region composed of several biomes, we decide to divide the study region into five subregions delimited by polygons (Fig. S7). The polygons were located in a latitudinal gradient covering relatively homogeneous areas within each polygon based on the authors' knowledge of the study area.

We used a supervised classification for each polygon for the years 2000 and 2020. We used a total of 350 points per polygon to perform the classification form which we used 85% for training and 15% for test. We used a set of geographic products as ancillary data to confirm that both, training and test points correspond to one of three classes: forest, no- forest, and water bodies. This ancillary data was the official land use land cover products of Mexican government edition three and six (INEGI, 2017, 2003), the Madmex land cover product for the year 2018, and the Google Earth desktop app. We do not distinguish between different types of forest (e.g. tropical rainforest, tropical dry-forest, temperate forest, etc.) nor succession stage (old forest or secondary forest) since our main objective was to assess changes in forest cover in general. We used the Random Forest classification algorithm because, in a prospective exercise, this yields the best performance in comparison to Classification and Regression Trees (CART), Support Vector Machines (SVM), and the Hansen cover product (Hansen et al., 2013).



Figure S7: The five polygons in which the study region was divided.

We calculated the classification accuracy for each polygon for years 2000 and 2020 (Table S5) using the caret package in R studio. We also assessed the classification accuracy inside the eighteen studied biosphere reserves (Table S6). To do that, we used 200 random points seeded inside each reserve. We used ancillary data to verify the actual land cover in the location of each point for the year 2020.

Table S5: Classification accuracy indicators for training and test data for the five polygons in which the study region was divided for the years 2000 and 2020.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Polygon | Year | Train overall accuracy | Train kappa | Test overall accuracy | Test kappa |
| p01 | 2000 | 0.9930 | 0.9770 | 0.9600 | 0.8344 |
| p02 | 2000 | 0.9900 | 0.9790 | 0.9400 | 0.8800 |
| p03 | 2000 | 0.9800 | 0.9599 | 0.9400 | 0.8682 |
| p04 | 2000 | 0.9130 | 0.8264 | 0.9600 | 0.9168 |
| p05 | 2000 | 0.9300 | 0.8598 | 0.9200 | 0.8379 |
| **Average** |  | **0.9612** | **0.9204** | **0.9440** | **0.8675** |
|  |  |  |  |  |  |
| p01 | 2020 | 0.9967 | 0.9892 | 0.9800 | 0.9117 |
| p02 | 2020 | 0.9900 | 0.9791 | 0.9000 | 0.7899 |
| p03 | 2020 | 0.9967 | 0.9933 | 0.9800 | 0.9596 |
| p04 | 2020 | 0.9700 | 0.9397 | 0.9200 | 0.8390 |
| p05 | 2020 | 0.9091 | 0.8197 | 0.9400 | 0.8796 |
| **Average** |   | **0.9725** | **0.9442** | **0.9440** | **0.8760** |

Table S6: Classification accuracy indicators for 18 biosphere reserves units. The classification corresponds to the year 2020.

|  |  |  |  |
| --- | --- | --- | --- |
| No | Reserve | Overall accuracy | Kappa |
| 1 | Barranca de Metztitlán | 0.8800 | 0.7200 |
| 2 | Calakmul | 0.9750 | 0.7869 |
| 3 | Chamela-Cuixmala | 0.9100 | 0.7162 |
| 4 | El Triunfo | 0.9450 | 0.8084 |
| 5 | La Encrucijada | 0.7990 | 0.5986 |
| 6 | La Sepultura | 0.8250 | 0.5750 |
| 7 | Lacandona complex | 0.9300 | 0.7122 |
| 8 | Los Tuxtlas | 0.9450 | 0.8684 |
| 9 | LPRC complex | 0.9343 | 0.7487 |
| 10 | Mariposa Monarca | 0.9450 | 0.8078 |
| 11 | Ría Lagartos | 0.9594 | 0.9114 |
| 12 | Selva El Ocote | 0.9250 | 0.7727 |
| 13 | Sian Ka'an | 0.9444 | 0.7263 |
| 14 | Sierra de Huautla | 0.9700 | 0.8961 |
| 15 | Sierra de Manantlán | 0.9200 | 0.6396 |
| 16 | Sierra del Abra Tanchipa | 0.9350 | 0.4118 |
| 17 | Sierra Gorda | 0.8900 | 0.5908 |
| 18 | Tehuacán-Cuicatlán | 0.8000 | 0.5035 |
|   | **Average** | **0.9129** | **0.7108** |

**Covariates calculation**

We used the distance to cities, distance to roads, and agriculture suitability index as covariates for matching analysis. We calculated the distance to cities, defined as localities with a population of 15,000 or higher, through a geographic information system (GIS). To do that, we used vectorial data of Mexican localities for the year 2000 (CONABIO, 2002). We generated a raster dataset of 1 km of cell size that contains the information.

To calculate the distance to roads we used vectorial data of the national roads network (INEGI, 2018). We assessed the distance to roads through the GIS and generated a raster dataset of 1 km of cell size. We calculated the agriculture suitability index using data of climate, soil, and orography. As climate variables, we considered the mean annual temperature, temperature annual range, mean annual precipitation, and precipitation of the driest quarter (Fick and Hijmans, 2017). As soil variables, we used the concentration of Ca, Na, organic carbon, and pH (INEGI, 2013a). And as orographic variables, we use elevation and slope (INEGI, 2013b). We obtained data on the presence or absence of agricultural lands from the Mad-Mex land use/cover classification. Mad-Mex provides information on 17 land cover classes at 30m of resolution. We reclassified these classes in a binary raster of agriculture/non-agriculture.



Figure S8: a) Predicted response curves of the explanatory variables as predictors of agriculture suitability index according to binomial model. b) Relative importance of each explanatory variable determined by Pearson correlation permutation.

To calculate the agriculture suitability index, we used binomial generalized linear models. As response variable, we used presences-absence of agriculture and we trained the model with all the set of climate, soil, and, orographic variables previously described. We used 1600 random points isolated at a distance of at least 1 km each other and seeded in the entire Mexican territory to sample the variables. Through this approach, we described the combination of conditions where agriculture activities were developing in the country with a low level of spatial autocorrelation (Moran’s I= 0.0012, p=0.79). To select a set of candidate explanatory variables, we tested for their collinearity and their relative importance through correlation and area under the curve metrics. We selected the most parsimonious model based on AIC criteria. Finally, the selected explanatory variables were slope, organic carbon, mean annual temperature, and mean annual precipitation (Fig. S8). We predicted agriculture suitability for the entire Mexican territory. We evaluated the accuracy of the model through the ROC curve (Fig. S9). All the analysis was performed using the *sdm* package of R (Naimi and Araújo, 2016).



Figure S9: Receiver operating curve (ROC) of the agriculture suitability model. The area under the curve (AUC) value is shown in the bottom right part of the plot. The magnitude of AUC indicates a medium level of the model prediction.

**Matching analysis methodology**

Two different strategies were used to carry out the matching analysis. On the one hand, for the data on forest loss and regeneration, a total of one million sample points uniformly seeded in the Mesoamerican region were used. Each point had the information about the covariates, whether there was forest loss, forest regrowth, or no cover change in the specific location of the point during the period 2000 2020, and on one of the two conditions: the treatment (i.e. reserves) and the control (unprotected areas). On the other hand, a total of 1500 microlandscapes were randomly seeded in the Mesoamerican region. Each landscape was defined by a circle of 3km of radius which is the maximum distance that perfectly fits inside the smallest reserve (Fig. S2, in Appendix 1). This approach is similar to the used in previous studies to evaluate fragmentation through matching analysis (Sims 2014). We quantified the number of forest patches in the year 2000 (NP2000) and in the year 2020 (NP2020) for each landscape using the *landscapemetrics* package of R (Hesselbarth et al., 2019), then we calculated the forest fragmentation rate (FFR) as follows:

$$FFR=\left(1-\left(1-\frac{NP\_{2020}-NP\_{2000}}{NP\_{2000}}\right)^{^{1}/\_{20}}\right)\*100$$

We also calculated the mean value of the covariates for each landscape. Finally, each landscape was assigned to control or treatment according to whether they were located inside a reserve or in unprotected zones. We avoided that the landscapes had an intersection of protected and unprotected zones by moving them a few meters. Data with information from any other type of PAs that are not biosphere reserves were not included in the analysis.

To achieve covariate balance, we used the nearest neighbor covariate matching algorithm with sampling replacement for forest loss and forest regrowth data, and the genetic covariate matching algorithm with sampling replacement and 150 generations for fragmentation data. The propensity score was estimated using binomial error distribution with logit function where the response variable was the treatment (protected or unprotected). After matching, all standardized mean differences for the covariates were below 0.1 which indicates a good balance between control and treatment samples. The matching analysis was performed using the *MatchIt* package (Ho et al., 2011) of R(R Core Team, 2021).

**Indicators of underlying drivers**

The population density was calculated by dividing the number of people in a municipality in the year 2000 by the municipality area. The population growth rate was calculated as r = ln(N2020/N2000)/20, which corresponds to the intrinsic population increase rate per year. Here, N2000 is the total population from all municipalities that intersect a given reserve in 2000, N2020 is the total population from the same municipalities ten years later (2020). Rural settlement density was estimated from vector data of Mexican rural localities in 2000 (CONABIO, 2002). We calculated Kernel density in QGIS, with a searching radius of 5 km and a cell size of 100 m. We calculated line density in QGIS using a searching radius of 5km and a cell size of 100m and we determine the mean value of road density by each reserve. Government subsidies for agriculture were calculated from PROAGRO program data (SAGARPA, 2018), which lists the amount of money given to each municipality per agriculture cycle (two cycles per year) for the period 2013-2018. We then calculated the total amount of money given for this period per municipality per unit area (Mexican pesos invested per square kilometer). To estimate non-farm occupation, we calculated the proportion of the population in a municipality working in the industrial or services sector. To calculate the distance to major cities, which here was used as a proxy to access markets, we used vector data of localities in the year 2000 with a population >15,000 (CONABIO, 2002). Then we calculate vector point distance in QGIS as raster data at a resolution of 100m. Finally, we calculate the mean distance to major cities for each reserve. Indicators such as marginalization index, human development index, and unemployment index were obtained from the different sources in a format that did not need any calculation.

Since most socioeconomic indicators were gathered from the municipality scale, we calculated the mean value of those that influence directly each reserve. To do that, we defined a minimum intersection area threshold of 10%. Thus, a municipality with an area higher than 10% covered by a reserve was considered to have a direct influence (Table S7). In this form, from the original 297 municipalities that intersect at some level any reserve, we obtained 220 that meet the 10% coverage threshold.

We calculated Pearson correlation coefficients of indicators of underlying drivers of forest spatial changes (forest loss, fragmentation, and regrowth). We discarded variables highly correlated (correlation value higher than 0.5). We also tested for multicollinearity through the variance inflation factor (VIF) and we exclude variables with values higher than 2. Thus, we discarded the subsidies for agriculture, unemployment rate, human development index, and marginalization index (Figure S3, Appendix 1). In this way, for our analysis, we selected unemployment rate, rural density, population density, population growth, non-farm occupation, and distance to major cities.

Table S7: The number of municipalities included to calculate the mean value of socioeconomic indicators per reserve.

|  |  |  |
| --- | --- | --- |
| no | Reserves | Municipalities |
| 1 | Barranca de Metztitlán | 17 |
| 2 | Calakmul | 4 |
| 3 | Chamela-Cuixmala | 1 |
| 4 | El Triunfo | 13 |
| 5 | La Sepultura | 8 |
| 6 | Lacandona | 4 |
| 7 | Los Tuxtlas | 9 |
| 8 | LPRC | 8 |
| 9 | Mariposa Monarca | 15 |
| 10 | Ría Lagartos | 4 |
| 11 | Selva El Ocote | 5 |
| 12 | Sian Ka'an | 2 |
| 13 | Sierra de Huautla | 13 |
| 14 | Sierra de Manantlán | 11 |
| 15 | Sierra del Abra Tanchipa | 3 |
| 16 | Sierra Gorda | 18 |
| 17 | Tehuacán-Cuicatlán | 85 |

|  |
| --- |
| Table S8: Socioeconomic indicators of underlying drivers of forest spatial changes.  |
| No | Indicator | Type of driver | Source | Period | Scale | Resolution | Range | Units | Brief description |
| 1 | Marginalization | Economic | CONABIO (2006) with information of CONAPO | 2000 | Municipality | - | 0 to 100 | Dimensionless | It is an indicator that assesses the intensity of privations suffered by the population. It takes into account education, households, population distribution, and income. The higher the value the more the deprivation. |
| 2 | Human development index (HDI) | Politic | CONABIO (2014) with information of PNUD Mexico | 2000 | Municipality | - | 0 to 1 | Dimensionless | It is an indicator of the effectiveness of the state to provide adequate conditions for the proper development of people's lives, taking into account income, health, and education. |
| 3 | Unemployment rate | Economic | Derived from CONABIO (2010a) with information of INEGI | 2000 | Municipality | - | 0 to100 | Percentage | It is the percentage of the economically active population with no job |
| 4 | Non-farm occupation | Economic | Derived from CONABIO (2010a) with information of INEGI | 2000 | Municipality | - | 0 to 1 | Proportion | The proportion of the population of a municipality that works in a non-farm sector (i.e. services or industrial). |
| 5 | Population growth rate | Demographic | CONABIO (2010b, 2010c) with information from INEGI | 2000-2020 | Municipality | - | -∞ to ∞ | Individuals/year | It is an indicator of the increase or decrease of the population in the period studied. |
| 6 | Population density | Demographic | Derived from CONABIO (2010b) with information of INEGI | 2000 | Municipality | - | 0 to ∞ | Individuals/km2 | Number of individuals per km2 |
| 7 | Government subsidies for agriculture | Politic | Derived from SAGARPA (2018) | 2013-2018 | Municipality | - | 0 to ∞ | Pesos/km2 | Amount of money per km2 invested in the municipality in the period. |
| 8 | Distance to cities | Economic | Derived from CONABIO (2002) with information of INEGI | 2000 | Pixel | 100m | 0 to ∞ | km | Distance to localities with 15,000 people or more. Distance to cities where there is a major market for trade in goods and services |
| 9 | Rural settlement density | Demographic | Derived from CONABIO (2002) with information of INEGI | 2000 | Pixel | 100m | 0 to ∞ | Settlements/km2 | Number of settlements per km2 |

**Supplementary references**

CONABIO, 2014. Índice de desarrollo humano por municipio, 2000 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/ (accessed 4.1.20).

CONABIO, 2010a. Empleo en México por municipio, 2000 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/ (accessed 4.1.10).

CONABIO, 2010b. Distribución de la población en México por municipio, 2000 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/ (accessed 4.1.00).

CONABIO, 2010c. Distribución de la población en México por municipio, 2010 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/ (accessed 4.1.20).

CONABIO, 2006. Grados de marginación a nivel localidad, 2000 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/ (accessed 4.1.20).

CONABIO, 2002. Localidades de la República Mexicana, 2000 [WWW Document]. Catálogo metadatos geográficos. Com. Nac. para el Conoc. y Uso la Biodivers. URL http://www.conabio.gob.mx/informacion/gis/

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

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., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science (80-. ). 342, 850–853. https://doi.org/10.1126/science.1244693

Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J., 2019. landscapemetrics: an open-source R tool to calculate landscape metrics. Ecography (Cop.). 42, 1648–1657. https://doi.org/10.1111/ecog.04617

Ho, D.E., King, G., Stuart, E.A., Imai, K., 2011. MatchIt : Nonparametric preprocessing for. J. Stat. Softw. 42, 1–28. https://doi.org/10.18637/jss.v042.i08

INEGI, 2018. Red nacional de caminos RCN. 2018 [WWW Document]. Inst. Nac. Estadística y Geogr. URL https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463674641 (accessed 11.15.19).

INEGI, 2017. Conjunto de datos vectoriales de la carta de uso del suelo y vegetación serie VI. Conjunto Nacional [WWW Document]. Inst. Nac. Estadística y Geogr. URL https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463598459 (accessed 9.1.17).

INEGI, 2013a. Conjunto de datos de perfiles de suelos. Escala 1:250 000 Serie II. [WWW Document]. Inst. Nac. Estadística y Geogr. URL https://www.inegi.org.mx/temas/mapas/edafologia/ (accessed 5.20.19).

INEGI, 2013b. Continúo de elevaciones mexicano 3.0 [WWW Document]. Inst. Nac. Estadística y Geogr. URL http://www.beta.inegi.org.mx/app/geo2/elevacionesmex/ (accessed 1.5.19).

INEGI, 2003. Conjunto de datos vectoriales de la carta de uso del suelo y vegetación serie II. Continuo Nacional [WWW Document]. Inst. Nac. Estadística y Geogr. URL https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825267865 (accessed 9.1.17).

Jost, L., 2006. Entropy and diversity. Oikos 113, 363–375.

Naimi, B., Araújo, M.B., 2016. Sdm: A reproducible and extensible R platform for species distribution modelling. Ecography (Cop.). 39, 368–375. https://doi.org/10.1111/ecog.01881

R Core Team, 2021. R: A Language and environment for statistical computing.

SAGARPA, 2018. Listado de beneficiarios PROAGRO productivo [WWW Document]. Secr. Agric. Ganad. Desarro. Rural. Pesca y Aliment. URL http://www.sagarpa.mx/agricultura/Programas/proagro/Beneficiarios/Paginas/Beneficiarios.aspx (accessed 4.5.19).