Massachusetts CAPS Input Data Layers – February 2021

Source: https://umasscaps.org

This document describes the source input data to CAPS, with a brief listing of major errors and limitations and of modifications we made to the data listed for each source. All data are the most recent available as of summer 2020. Most of these data are available from MassGIS (https://www.mass.gov/orgs/massgis-bureau-of-geographic-information).

MassGIS 2016 land cover/Land use (2019) - This is the source for most developed land and and natural types in the land cover. Natural communities include forest, shrublands, and various wetlands. MassGIS created this layer by combining NOAA's Coastal Change Analysis Program (C-CAP) land cover data with MassGIS' Standardized "Level 3" Parcels layer. The original from MassGIS is a high-resolution polygon coverage with separate values for land cover and land use. We reprojected the data to Mass State Plane, and converted it to a 30 m grid to align with the rest of our data. After extensive evaluation, we crosswalked the combined layer into our raw land cover, which captures developed and natural land cover types that are meaningful for the CAPS analysis. We also derived an impervious surface layer from this source. We combined our raw land cover with several other layers to make capsland, our final land cover: roofprints (representing all buildings), roads, railroads, streams, vernal pools, power lines (as shrubland), dams, and road-stream crossings.

Notes

- Wetlands now are mapped in the 2016 land cover by NOAA's C-CAP (2016). In previous versions of CAPS, our wetlands came from DEP wetlands (2005), which were photo-interpreted and of high quality, but outdated, based on aerial photos from as long ago as the 1990s. As a result, many beaver-created and impacted wetlands were mismapped. The new NOAA wetlands have less thematic richness, not distinguishing between shallow and deep marsh, and dropping the bog class (bogs are usually mapped as marsh, shrub swamp, or sometimes, aquatic bed). We didn't think the new "aquatic bed" class was terribly reliable nor meaningful for our purposes, so we reclassified aquatic bed to lake or pond. Cranberry bogs are no longer separately mapped; in this analysis, these are usually mapped as cropland. We brought in vernal pools from our 2011 land cover; these are based on Natural Heritage's potential vernal pool layer.
- As in previous versions, we split open water into lotic and lentic based on a model that used the shape of polygons to distinguish rivers from lakes and ponds, followed by thorough hand-checking and editing. We further split lentic into lakes and ponds based on the size of the waterbody (ponds are < 8 ha). This was based on a logistic regression of sizes of lakes and ponds in the National Wetlands Inventory, because NWI distinguishes between lakes and ponds, whereas 2016 land cover depicts all open water as one class. Rivers and streams are mapped by size and gradient, as in previous versions (see below). We used our 2011 classification of ocean (with misalignment correction) to remap the generic open water class to ocean.

- Sea cliff, rocky intertidal, and vegetated dune are no longer mapped.
- In conjunction with roofprints, we were able to split developed land into buildings, pavement (mapped as impervious), and the new class, developed open space.
- The new grasslands class were a mess. In mainland Massachusetts, where there are very few natural grasslands, the majority of areas mapped as grassland were really developed open space: baseball fields, pastures, mowed lawns surrounding sewage plants, solar farms, gravel pits, and cranberry bogs. Some heavily-cut forests were also mapped as grasslands. On the Cape and Islands, natural grasslands were usually correctly mapped, but unfortunately, on the order of half of areas mapped as grasslands were really developed open space. Given the importance of grasslands in areas where they occur, we kept grasslands on the Cape and Islands, despite the errors of commission. In mainland Massachusetts, we converted all mapped grasslands to developed open space. In addition, large areas of forest on the Prescott Peninsula in the Quabbin were inexplicably mapped as grassland (these areas were far larger than any areas of forest harvest, and have not been heavily cut in recent decades). We converted all bogus grasslands on the Prescott Peninsula to forest.
- Coastal dunes were typically mapped in MassGIS 2016 land cover as developed open space or barren land. We rectified dunes by taking areas mapped by the 2005 DEP wetlands as coastal dune or vegetated dune and mapped by 2016 land cover as developed open space or barren land and converting them to coastal dune (with some effort to correct for misalignment). Vegetated dune areas are typically mapped as grassland.
- We remapped barren land that falls on or near coastal beaches, tidal flats, or ocean in the 2005 DEP wetlands to beach or mudflat.
- Powerlines, which were previously mapped as "powerline shrubland," were typically mapped in MassGIS 2016 land use as shrubland, pasture, or grassland, but also as forest or other classes. We changed areas under 2005 powerlines that were mapped by 2016 land cover as pasture, grassland, developed open space, barren land, or forest to shrubland.
- The new barren land class is something of a garbage can. Although it no doubt makes sense from a spectral mapping standpoint, it doesn't in an ecological sense. It represents a mix of gravel pits, talus slopes, coastal beaches, sea cliffs, lakeshores, and anthropogenic edges. After pulling out dunes, beaches and powerlines (see above), we mapped remaining areas of barren land as bare land, and excluded it from estimates of ecological integrity, as these areas are primarily anthropogenic.
- Nomans Land (an island southeast of Martha's Vinyard) was totally omitted in the MassGIS data. We brought it in from the 2011 Massachusetts CAPS land cover.
- There are several known errors in the landcover due to errors in the 2016 land cover. The only errors we specifically fixed were several large areas of forest mapped as grassland on the Prescott Peninsula, as these were so egregious. Notable errors include some large areas of recently managed pitch pine and scrub-oak in the Montague Sandplains being mapped as developed open space, heavily cut forests mapped as

developed open space, and small areas of trees in the middle of urbanized areas being mapped as forests (the original C-CAP data sometimes maps individual trees as forest!).

Roofprints – We used the MassGIS 2016 building structures (2019) to represent buildings. Combined with the MassGIS land cover representation of impervious surfaces (which we interpret as pavement in cells without roofprints) and developed open space (managed semi-natural space such as lawns, parks, and cemeteries), roofprints give us a more finely detailed and accurate representation of various types of development. We combined roofprints with MassGIS land use to distinguish six types of buildings: commercial, industrial, agricultural, residential, recreational, and public buildings.

Note that the MassGIS roofprints seem to be of higher quality than the Microsoft Bing roofprints we used for the region-wide DSL project, which itself is of fairly high quality—certainly far better than any representation of development we've had access to in the past.

Potential vernal pools – We used photo-interpreted Potential Vernal Pools from MassWildlife's Natural Heritage and Endangered Species Program.

- Potential vernal pools that fell within a terrestrial type were treated as a single pixel pool (30 m × 30 m). When a potential vernal pool fell within a wetland mapped by DEP, we retained DEP's classification.
- Because of the inherent difficulty of identifying vernal pools from aerial photography, this layer contains many errors of commission and omission. Because there is no other data source for this important community (certified vernal pools are still quite limited and highly biased by search effort), we used these data with caveats.
- We moved vernal pools that fell in the same cell as a road over one cell to fall alongside the road in our land use. We used an algorithm that looked at the vector data to move the pool to the correct side of the road.

MassGIS networked hydro centerlines, NHD stream network – We used the MassGIS networked stream centerlines for the mainland, and filled in the Cape and islands with edited versions of NHD centerlines.

- We edited these data to repair a significant number of breaks in the network, as the CAPS watershed metrics require a connected network.
- We deleted the dense (and meaningless) network of channels in cranberry bogs, instead connecting streams flowing through bogs with straight lines passing through the bogs. These dense channels made it impossible to represent flow in a 30 m grid.
- We deleted the ditches in salt marshes for similar reasons. Streams that flowed into DEP salt marshes were retained, and any stream that originated within a salt marsh was deleted.
- We extended stream mouths all the way to the ocean

- We added stream centerlines to our land cover grid in areas that were mapped as uplands to represent smaller (1st and 2nd order) stream communities
- We burned stream centerlines into the flow grid to force streams and rivers to prevent small DEM errors from misdirecting streams and rivers
- Because stream centerlines were digitized at varying densities, resulting in bias that affected our aquatic connectedness metric, we dropped all streams with a watershed of less than 30 ha. This has the effect of removing parts of the smaller headwater streams throughout, while making stream density more consistent.

MassGIS 5 m Digital Elevation Model – The DEM is the basis of several of our terrain-based settings variables, and of the flow used for our watershed metrics. We used the 5 m DEM because its accuracy, consistency, and overall quality was much higher than the older 30 m DEM and the DEM from the National Elevation Dataset (NED). We used the DEM to create a flow direction grid, the source of flow accumulation and $CaCO_3$ settings variables, and watershed metrics. The DEM was also used to model tides and tidal restrictions, solar exposure, and the slope and gradient settings variables.

- We sampled this DEM up to 30 m for all analyses
- For flow modeling (flow-based settings variables and watershed metrics), we filled depressions in the DEM

Flow direction – The D8 (single-direction) flow direction grid was derived from the depressionless 30 m DEM. We then burned stream centerlines into the flow grid to ensure that stream and riverbeds are represented correctly. This was an iterative processes that entailed finding loops introduced by errors in stream centerlines, correcting them, and repeating the process. Flow direction is used for watershed metrics, and also for the $CaCO_3$ content settings variable.

Flow accumulation – We built a FD8 (multiple-direction) flow accumulation grid from the flow grid and DEM. This process allows a cell to flow to multiple downslope cells, giving much more realistic flow patterns in mid-slopes. Flow accumulation is used for the settings variables wetness and flow volume.

- We estimated the watershed area of streams flowing into Massachusetts from the 2009 National Elevation Dataset (NED) flow accumulation grid.

Aquatic resistance – We modified the approach of Randhir et al. (2001) to build a time-of-travel grid for each cell in the project area, based on land cover, slope, flow, and stream gradient. This grid was used to define the influence area within the watershed of each point for our watershed metrics.

Dams – Dams are from Northeast Aquatic Connectivity Assessment Project (NACAP) version 2 (2017). They were processed to fall on our DSL stream centerlines by Ethan Plunkett, and adjusted to match the Massachusetts stream centerlines (which are very similar).

Protected open space – We selected permanently protected open space (lev_prot = "P") from the February 2020 MassGIS protected and recreational open space layer. These data are used to adjust road traffic rates in parks and state forests.

Roads and road traffic – Roads are from the 2018 MassDOT 1:5000 Roads layer (via MassGIS). Roads were reclassified into five types (expressway, primary highway, secondary highway, light duty road, and unpaved road) based on original road classes as well as surface type for unpaved roads.

Traffic rates are a mess in the 2018 roads layer. Traffic rates were measured in the field for many roads, especially larger roads. For unmeasured roads, MassDOT assigned arbitrary fixed traffic rates, varying nonsensically by region; e.g., 1000 in much of western Massachusetts, 898 in Springfield, 1069 in Pittsfield. The previous traffic rate model used by MassDOT (which seemed pretty good) has been dropped. As traffic rates are important to both the traffic and connectedness (and to a lesser extent, similarity) metrics, the data as delivered weren't acceptable for CAPS, so we built a suite of road traffic models ourselves.

We treated the traffic rates differently for each road class:

- **Expressways** (class 1) – as most expressways were measured in the field, only a handful of segments were missing traffic rates. We filled these in by hand from either the opposite lane or adjacent road segments.
- **Primary highways** (class 2) – we fit primary highways using a quantile regression to fit median traffic rates from a number of variables including kernels of development at various scales, kernels of road density at varying scales, kernels of agriculture and of protected open space, and interpolations of class 2 traffic rates. We used AIC to select the best model, which included two interpolations of class 2 traffic for non-missing road segments (an inverse distance weighted interpolation and a kriged interpolation). We stratified our sample to prevent overrepresentation of more common traffic rates. We assessed all fits with a large holdout sample, and got an excellent fit, with a log-log R^2 of 0.8998. As class 2 highways are relatively sparse on the landscape and many segments were surveyed, to a great extent the interpolations had the effect of correctly filling in missing traffic rates from nearby surveyed samples.
- **Secondary highways** (class 3) – we fit secondary highways using the same approach as for primary highways. The best-fitting model used inverse distance weighted interpolation of class 3 and class 4 traffic rates, a kriged interpolation of class 4 traffic rates, and a kernel of class 3 road density. The fit was pretty good, with a log-log R^2 on the holdout set of 0.6322.
- **Light duty roads** (class 4) – the quantile regression model we used for class 2 and 3 roads performed poorly for light duty roads, so we ended up using a Random Forest model to fit

traffic rates from kernels of development, agriculture, open space, inverse-distance weighted class 2, 3, and 4 interpolations of traffic rates, kriged class 2, 3, and 4 traffic rates, and kernels of class 2 and 3 road density. The R^2 was 0.6135.

Unpaved roads (class 5) – vanishingly few unpaved roads had measured traffic rates, and many of these were obviously classification errors, so it was impossible to model traffic rates for unpaved roads from the sampled roads. Instead, we applied the model for light duty roads and multiplied estimated traffic rates by 20%. We of course had no basis to pick this discount, other than knowing intuitively that dirt roads typically have much lower traffic than paved roads. The results are of unknown accuracy, but are clearly far better than the absurdly high rates assigned regionally by MassGIS.

As in previous versions, we changed the traffic rate for all unpaved roads that run through permanently protected open space to 10. This fixes the wild overestimates of traffic on gated, discontinued, and little-used roads through parks and state forests such as Myles Standish, the Quabbin, and many other state forests and other large conservation areas.

Railroads – From the 2015 MassGIS trains layer. Railroads were mapped in three classes: railroad, abandoned railbed, and rail trail.

- We deleted linework where abandoned rails were shown underwater in the Quabbin, and where railroads run through major tunnels (the Hoosac tunnel)
- We integrated rail traffic into our traffic rates layer by assigning traffic rates for commuter, passenger, and freight lines, based on estimates of the number of cars for each railroad type, estimated average number of daily freight and commuter trains, and number of daily passenger trains from schedules, and expert team assignment of the relative impact of a train car to an automobile.
- We estimated the number of tracks in each rail line from GIS data for use in the terrestrial barriers settings variable.

Road-stream crossings – Bridge and culvert locations were estimated from the intersections of stream centerlines with road and railroad linework. For each crossing, we obtained aquatic and terrestrial passability scores from the North Atlantic Aquatic Connectivity Collaborative (NAACC). We used field-surveyed scores where available (scores were downloaded on July 26, 2020), and estimated scores elsewhere using a random forest model developed by Ethan Plunkett.

- Bridges and culverts on small, unmapped streams and unmapped roads are omitted.
- Bridges and culverts on streams with < 30 ha watersheds are omitted because we trimmed streams with very small watersheds to reduce bias from inconsistent effort by USGS when they digitized streams.
- Crossing scores are based on a modeling approach with rather wide confidence intervals; furthermore, models of target scores are based on expert opinion rather than empirical passability data.

- Aquatic passability is used for the aquatic barriers settings variable.
- Terrestrial passability is used in the connectedness metric to allow connectivity under roads and railroads at stream crossings.

Tidal regime – Tidal regime is estimated from a logistic regression of salt marshes (from 2015 MassGIS land cover) vs. uplands from the DEM and interpolated tide range from 120 NOAA tide stations. This gives a grid depicting the expected tidal influence at each point. Tidal regime is used as a settings variable, and also as an input to the tidal restrictions metric.

Tidal restrictions – Potential tidal restrictions are modeled at road-stream crossings (the intersection of stream centerlines with road and railroad linework) in the coastal area. In previous versions, we modeled the severity of tidal restrictions based on 75 measured tidal restrictions from field work done by CZM and DEP using a regression of the ratio of the area of expected salt marsh above each restriction (areas where the tidal regime suggest salt marshes) to the area of salt marsh mapped by DEP above each restriction ($P < 0.001$; $r^2 = 0.41$). However, many of these field-measured restrictions were at remediation sites where culverts have been upgraded in the past 15 or 20 years; as a result, given new landcover, our regression no longer fits. We now use the same approach to modeling tidal restrictions as used by the DSL project, described here:
http://landeco.umass.edu/web/lcc/dsl/metrics/DSL_documentation_tidal_restrictions.pdf.
Note that many restrictions occur where there are no roads or railroads, for instance at tide gates; we were unable to capture such restrictions

Imperviousness – Impervious surfaces are from the 2016 MassGIS land cover/land use. We converted the impervious, roofprint, and roads shapefiles to 1 m rasters, after buffering roads by class to represent typical road widths. We combined these layers, and converted them to percent impervious within each 30 m cell.

Soils – Soil depth, pH, and texture are from NRCS digital soil maps. We used 1:25,000 SSURGO soils data where possible. In Franklin and Plymouth counties, SSURGO pH data were spotty, so we layered SSURGO pH on top of the more complete but more generalized STATSGO (1:250,000) values.

- Soil texture was classified on an ordinal scale of 1-6, where 1 is organic, and 2-6 range from fine to coarse textured. Values were lumped based on texture classes such as “silt loam” or “very fine sandy loam.” Soil texture was not supplied for open water or urban areas.
- Soil pH was based on the representative pH for each soil type. pH values are fairly coarse, with missing values for urban areas, open water, and many other areas. Soil pH for Franklin and Plymouth counties were a mixture of low-resolution STATSGO and high-resolution but incomplete SSURGO data.
- Soil depth is the expected depth to bedrock, dense, or cemented layers. We log-transformed soil depth for the soildepth settings variable. Soil depth is missing for open

water and is set to zero for some mountainous areas and other apparently arbitrary areas.

Calcium – We used the geology field from TNC's Ecological Land Units for calcareous and moderately calcareous near-surface bedrock. (The original source for this dataset is USGS, available on MassGIS; TNC has reclassified lithology). Our $CaCO_3$ settings variable uses these values directly for terrestrial areas, and uses a flow accumulation model for wetlands and open water.

- Lithology is mapped at scales ranging from 1:100,000 to 1:500,000, so fine details and smaller inclusions are omitted, and spatial accuracy is poor.

Wind speed, wind power – Wind speed and power data are modeled by TrueWind Solutions LLC. Wind speed is available from MassGIS; we obtained wind power (in 16 cardinal directions) from the UMass Wind Energy Center. Original data are at 200 m resolution; we downsampled these data to 30 m by interpolation.

- Our wind exposure settings variable is based on wind speed at an elevation of 30 m.
- The wave exposure settings variable uses directional window power at 50 m combined with reach to estimate potential wave exposure along the coast.

Salinity – Our salinity settings variable has three classes:

- Saltwater: areas mapped in MassGIS land cover (2016) as ocean.
- Brackish: areas mapped in DEP wetlands (2005) as salt marsh, areas mapped as open water tidal, brackish, or salt pond (poly_code = 9) in DEP wetlands, and additional areas photo-interpreted as brackish for this project by Mike McHugh at MassDEP.
- Freshwater: anything that is not saltwater or brackish.

Public beaches – MassGIS's marine beaches. Used for the beach pedestrians metric.

Beach off-road vehicles – Area where off-road vehicles congregate and park on beaches were mapped by Nathalie Regis and Mike McHugh at MassDEP for this project. Areas of intensive ORV use on beaches were mapped based on information from DEP and CZM personnel as well as photointerpretation of beaches in the MassGIS DPH marine beaches layer. These data are used for the Beach ORVs metric.

Recreational beach parking lots – Parking lots for access to recreational beaches were photo-interpreted by Nathalie Regis and Mike McHugh at MassDEP for this project. All parking lots that appeared to serve recreational beaches mapped in the MassGIS marine beaches layer were delineated. Data were modified based on review by experts at MassDEP and CZM. These data are used for the beach pedestrians metric.

Salt marsh ditches – Ditches in salt marshes were photo-interpreted for this project by Nathalie Regis and Mike McHugh. This layer is used for the salt marsh ditching metric.

Coastal structures – Seawalls, jetties, groins, bulkheads, revetments, and breakwaters were originally obtained for most of the Massachusetts coast in field surveys by CZM. These surveys omitted some areas where access was not feasible. Omitted areas were completed based on photointerpretation by Nathalie Regis and Mike McHugh using orthophotos and oblique aerial photography. Coastal structures are used in the coastal structures metric.

Boat traffic – Marine and estuarine boat traffic were estimated from three sources: 2012 Automatic Identification Systems (AIS) data, 2007-8 Vessel Monitoring System (VMS) data, and 2010-2012 Recreational Boater Routes (RBR). These layers were combined in a model developed by Marc Carullo (CZM), and Michael McHugh and James Sprague (MassDEP).

Minimum winter temperature, growing season degree-days – Temperature data were obtained by downscaling modeled PRISM weather data via interpolation. Data are 30-year normals centered on 1985.

- The minimum winter temperature settings variable is the minimum of the coldest day in January or February.
- The growing season degree-days settings variable is based on the sum of monthly mean temperatures above a threshold of 10˚ C and below a threshold of 30˚ C.

Layers for hydrologic alterations and nutrient enrichment. Several datalayers were used for the three new stream metrics developed by Elizabeth Homa (Homa et al. 2013). Layers include: minimum annual temperature, mean annual temperature, annual precipitation, December precipitation, humidity (PRISM), MassGIS land use (2005), NLCD imperviousness (National Land Cover Dataset), water discharges (Massachusetts Sustainable Yield Estimator database, USGS), septic systems (1990 U.S. Census), percent sand (NRCS STATSGO soils, 1997), and dam storage (TNC dams layer). We used the original data the regressions models were built on rather than the newest data used elsewhere, as updating the data could have had unknown effects on the models, and some values (such as dam storage) are not available in our newest data.

Ecoregions – EPA ecoregions for Massachusetts are from MassGIS. We modified these data slightly to include all coastal cells. Ecoregions are used for IEI rescaled by ecoregions (IEI_E) and integrated IEI (IEI_I).

Watersheds – Major watersheds are from MassGIS. We modified these data slightly to include all coastal cells. Watersheds are used for IEI rescaled by watersheds (IEI_W) and integrated IEI (IEI_I).