Processing of INEGI USOSV Dataset for Use in Hydrologic Modeling

Theodore J. Bohn

**1. Introduction**

Our goal was to get land cover classifications for Mexico for multiple snapshots over the last several decades, for the purpose of studying the hydrologic impacts of land cover change. For this purpose, we chose to use the Uso del Suelo y Vegetaciòn (USOSV) land cover classification produced by the Instituto Nacional de Estadística y Geografía (INEGI), which has vector-format classifications derived from LANDSAT imagery for the years 1985, 1993, 2002, 2007, and 2011 (INEGI 2014). To be useful for hydrologic modeling, we needed them in raster format, geographic projection at 1/2400 degree resolution, from which we could aggregate to appropriate coarser resolutions. In addition, because USOSV used slightly different classification categories in different years, each of which used over 50 classes, we needed to group these classes into a single, smaller set of more general classes that was both consistent across snapshots. Finally, to facilitate land cover change studies spanning the US-Mexico border, we needed this simpler set of classes to be compatible with the scheme used for the National Land Cover Database (NLCD) (Homer et al. 2015). This document describes the processing steps used to accomplish these goals.

**2. Source Files**

We obtained ESRI-format shapefiles for the INEGI USOSV (INEGI 2014) product series 1-4 from INEGI personnel during a visit to their offices. Series 5 was downloaded directly from the web at <http://www.inegi.org.mx/geo/contenidos/recnat/usosuelo/>. Series 1-5 correspond to the years 1985, 1993, 2002, 2007, and 2011, respectively.

The data for series 5 consisted of a single shapefile for the entire country of Mexico, in Lambert conical projection. In contrast, the data for series 1-4 consisted of separate shapefiles for each of approximately 100 1 x 2 degree tiles in UTM projection, covering the entire country of Mexico (see Fig. 1). These tiles were organized by UTM zones, with each zone being 6 degrees wide in the East-West direction. Mexico spans from zone 11 to zone 16. These UTM zones were further subdivided in the North-South direction into 4-degree intervals (Mexico spans from “D” to “I”), leading to 4 X 6 degree rectangles. Thus, each rectangle is named by the latitude interval followed by the UTM zone, e.g., “F13”. Each rectangle is further subdivided into 12 1 x 2 degree tiles, numbered “01” through “12”. Thus, each 1 x 2 tile is named by the name of its parent rectangle followed by its individual number within the rectangle, e.g., “F1302”. To reduce the number of files to work with, we initially combined groups of 1 x 2 tiles into “meta-tiles” as shown in Fig. 1.

**3. Reprojection to Geographic Coordinates**

Each meta-tile was reprojected to geographic projection separately, because each UTM zone requires different reprojection parameters. In addition, different snapshots were referenced to different geographic datums: series 1-2 used the NAD 1927 datum, while series 3-5 used the ITRF 1992 datum (similar to WGS 1984). After reprojection, meta-tiles were merged into a single file covering all of Mexico, which was used in all subsequent processing.

**4. Regrouping the Classes**

The different series used slightly different conventions for their classifications, with series 1-2 and series 5 storing all class information in a single field of a single file (or set of files), and series 3-4 storing some secondary information (e.g., on agricultural sub-types) in a separate set of files. In addition, the class naming conventions changed from series 1-2 to series 3-5, although most classes remained the same.

Our regrouping consisted of mapping each INEGI class to the most appropriate class of the NLCD 2011 legend, as shown in Table 1. Class codes were read from the “clavefot” field of series 1-2; the “clavefot” field “v” files and the “cve\_g” field of “g” files for series 3-4; and the “cve\_union” field of series 5.

4.1. Secondary Vegetation Codes

In the USOSV classification scheme, additional detail regarding secondary vegetation within a polygon of a given class was conveyed via prepending or appending extra codes (which we have not listed in Table 1):

* VSa – veg. secundaria arbustiva – shrub secondary vegetation
* VSA – veg. secundaria arborea – tree secondary vegetation
* VSh – veg. secundaria herbacea – herbaceous secondary vegetation

In series 1, the “VS\*” codes are appended to the end as “/VS\*”. In series 2-4, some classes have the “VS\*” appended as in series 1, and some classes have the “VS\*” prepended as “VS\*/”. In series 5, the “VS\*” is only prepended, never appended.

Unfortunately, INEGI’s practice of prepending or appending these codes to the primary vegetation class code (and doing so inconsistently across snapshots) effectively created many more classes and complicated the mapping process. Secondary vegetation information ideally would have been stored as a separate attribute. Thus, one processing step consisted of parsing out these secondary codes and removing them from the primary class code.

4.2. Erosion Information Codes

In a similar fashion to secondary vegetation, in series 1 and 2 only, information about erosion was conveyed in the class code:

* Prepended with “E-“ – some erosion present
* No “E-“ – no appreciable erosion

These codes also needed to be removed via a processing step before the primary classes could be mapped to NLCD classes. Again, this information ideally would have been stored as a separate attribute. Indeed, from series 3 onwards, the erosion information was stored in a separate file and did not pose a problem for processing.

4.3. Shrubland Sub-Classes

Similarly to secondary vegetation and erosion, in series 1 and 2 only, shrublands (“matorral”) were subdivided into sub-classes (e.g., inerme, subinerme, espinoso, etc) which were included in the code for the class via appending “/M\_” where “\_” is some letter. These codes are:

* MB – subinerme
* MCA – cardonal
* MCH - chollal
* ME – espinoso
* MF - cirio
* MH - herbazal
* MI – inerme
* MR – crasi-rosulifolios
* MZ – izotal

Again, these shrubland sub-class codes were split from the primary class codes as a processing step before mapping to NLCD classes. Again, this information ideally would have been stored as a separate attribute. Again, from series 3 onward, this information was stored in separate files and did not pose a problem for processing.

4.4. Changes in Primary Vegetation Classes Across Snapshots

The classifications of series 1 and 2 differed slightly from series 3-5. In Table 1, classes highlighted in yellow are present in series 1 and 2 but not later. Classes highlighted in green are present in series 3 onwards but not 1 or 2.

4.5. Agricultural Classes

Agricultural class code conventions changed back and forth across snapshots. In series 1 and 2, class codes reflected the presence/absence of annual (A), semi-permanent (S), and permanent (P) crops and ordered them in term of abundance (primary, secondary, tertiary). Thus, there were different classes for every possible combination and order of these crop types. In series 3 and 4, all agricultural classes and all anthropogenic pasture classes were denoted by the code “IAPF”; detailed agricultural and pastoral information was stored in the “g” files. In series 5, class codes again contained “A”, “S”, and “P” modifiers, but only up to two of these codes were included, and the modifiers only indicated presence/absence so that their order never varied. Again, our processing scripts removed these modifiers where present, before mapping to NLCD classes. These agricultural sub-types are not shown in Table 1.

**5. Processing Steps and Scripts**

All scripts and tools used for processing are stored in the “scripts” folder (except for the text editor step).

Processing steps were as follows (folder names are relative to location under “SERIE\_X” folder, where X = I, II, III, IV, or V):

1. For series 1-4, reproject the meta-tiles (shown in Figure 1) into geographic coordinates and WGS 1984 datum. This required 2 steps: 1) define a projection (using .prj files from series 3) via the script defproj\_inegi\_tiles.py; 2) reproject to geographic, via the script *reproj\_inegi\_tiles\_geo.py*. For series 3-4, these two operations were performed on both the “v” files and the “g” files. Input folder: “metatiles”; output folder = “geo”.
2. For series 1-4, merge the 11 meta-tiles into a single file for the entire country. Used arcpy script *merge\_inegi\_metatiles.py*. For series 3-4, these two operations were performed on both the “v” files and the “g” files. The output filenames will be of the form “usv250sXY.shp”, where X = 1, 2, 3, or 4 (equal to the series number) and Y = blank or “g” for input types of “v” or “g”, respectively. Input folder = “geo”; output folder = “entire”.
3. For series 5, reproject the original file from Lambert conical to geographic and WGS 1984 datum. Simply did this in ArcMap without running a script. Input file = “entire/usv250s5\_union.shp”; output file = “geo/usv250s5.shp”.
4. For all series, add and modify fields as necessary to end up with a consistent NLCD classification in the “ClassName” and “ClassNum” fields. Details: For series 1-4, add a “CVE\_UNION” field containing an INEGI code matching series 5’s CVE\_UNION field. For series 1-2, this primarily entailed: (a) removing the “E-“ flag that denoted erosion, (b) removing flags that indicated shrub (matorral) sub-types, and (c) swapping the order of the primary INEGI codes and the “VSa”, “VSA”, “VSh” qualifiers. Then, add “ClassName” and “ClassNum” fields to store the NLCD classes assigned to each shape. Translation between INEGI codes and the NLCD classification scheme is shown in Table 1. For series 3-4, this operation was performed on both the “v” files and the “g” files. Used arcpy scripts *add\_cve\_union.py* and *add\_cve\_union\_gfile.py*. Input files = “entire/usv250sXY.shp” (series 1-4) or “geo/usv250sX.shp” (series 5); output files = “cve\_union/usv250sXY.shp”.
5. For series 1-5, rasterize the shapefiles at 1/2400 (0.0004166666666667) degree resolution and output in ascii format. For series 3-4, this operation was performed on both the “v” files and the “g” files. Input folder = “cve\_union”; output folder = “raster” for the rasters, and “asc” for ascii files. Did this in ArcMap rather than with scripts.
6. For series 1-5, used the utility “vi” (any text editor would suffice for this) to replace the nodata value of -9999 with -1 everywhere in each ascii file. Because there are millions of pixels with nodata values, this change cuts the file size in half. Stored the outputs in the “asc.gmerge” directory.
7. For subsequent processing (combination with NLCD dataset, aggregation to 1/16 degree resolution, etc), clipped the files to 1x1 degree squares, stored in “asc.clip.1deg” directory. To accomplish this, used the C-language tool *gridclip*.

**4. Output**

The outputs are stored as follows:

SERIE\_X/ (where “X” is one of “I”, “II”, “III”, “IV”, or “V”)

(series 1-4 only):

metatiles/

Shapefiles merged into 11 meta-tiles in UTM projection. Series 1-2 use the NAD 1927 datum, while series 3-4 use the WGS 1984 datum. For series 3-4, there are two sets of files, “v” and “g” files. The “g” files contain the details for agricultural/pasture classes (which are all lumped into the “IAPF” class in the “v” files).

metatiles.geo/

Shapefiles in geographic projection with WGS 1984 datum, for meta-tiles. For series 3-4, there are normal files and “g” files; the “g” files contain details of agricultural and pasture classes, which are lumped into the “IAPF” class in the normal files.

entire/

Shapefiles for entire country, in geographic projection. Shapefile names must be “usv250sX.shp”, where X = 1, 2, 3, or 4 (the series number). For series 3-4, there are additional “g” files named “usv250sXg.shp”.

(series 5 only):

entire/

Shapefile for entire country in Lambert conical projection (the original file from the web). Shapefile name is usv250s5\_union.shp.

geo/

Shapefile for entire country in geographic projection with WGS 1984 datum. Shapefile name is usv250s5.shp. The “union” was stripped to make ArcGIS happy.

(all series):

cve\_union/

Shapefiles for entire country in geographic projection. Files have “CVE\_UNION” field with INEGI codes that are consistent across all series; and “ClassName” and “ClassNum” fields containing my simplified classification. For series 3-4, there are normal files and “g” files; the “g” files contain details of agricultural and pasture classes, which are lumped into the “IAPF” class in the normal files. Input filenames \*must\* all be “usv250sX.shp”, where X = 1, 2, 3, 4, or 5 (same as the series). For series 3-4, there are additional “g” files named “usv250sXg.shp”.

raster/

Rasterization of the files in cve\_union/, at 1/2400 (0.0004166666666667) degree resolution. For series 3-4, there are normal files and “g” files.

asc/

Ascii versions of raster files. For series 3-4, there are normal files and “g” files.

asc.gmerge/

Final ascii files, excluding the “g” files of series 3-4, with nodata value changed from -9999 to -1 to reduce file size. The name “gmerge” originally was intended to convey that the information from the “g” files was merged into the other files, but this is no longer applicable under the most recent version of the processing (in which all agricultural classes are grouped together as code 82).

YYYY/ (where YYYY = one of 1985, 1993, 2002, 2007, 2011)

asc.clip.1deg/

Ascii files that are the output of clipping the files from SERIE\_X/asc.gmerge/ into 1x1 degree squares.

Note: we have not included the rasters in this archive, in order to save space.

**References**

Homer, C. G., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.*, **81**, 345–354.

INEGI, 2014: Conjunto de datos vectoriales de Uso del Suelo y Vegetación, Escala 1:250 000, Serie V (Capa Unión). http://www.inegi.org.mx/geo/contenidos/recnat/usosuelo/ (Accessed October 1, 2015).

**Tables**

Table 1. Primary (i.e., not including “VSa”, “VSA”, and “VSh” modifiers) INEGI Codes, descriptions, and the NLCD class that each INEGI class was mapped to. Yellow indicates codes present in series 1-2 only. Green indicates codes present in series 3-5 only.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Code** | **Description** | **NLCD ClassName1** | **NLCD ClassNum1** | **Comment** |
| [R] | Area Agricola; Riego suspendido | Cultivated Crops | 82 | Present in s1,2 |
| ACUI | Acuicola | Cultivated Crops | 82 | Not present in s1,2 |
| ADV | Desprovisto de veg. | Barren Land | 31 | Not present in s1,2 |
| AH | Asentamientos humanos | Developed, low intensity | 22 | “human settlements”; seems to be applied to areas on the fringes of large cities, but not small towns, which are marked as “Zona urbana”; not present in s1,2 |
| BA | Bosque de oyamel | Needleleaf Forest | 42 | “Fir forest” |
| BB | Bosque cedro | Needleleaf Forest | 42 | “Cedar forest” |
| BC | Bosque cultivado | Broadleaf Forest | 41 |  |
| BG | Bosque de galeria | Broadleaf Forest | 41 | “Riparian forest”, Technically wetland, but seems to have lots of trees and bushes so probably would be classified as forest in remote-sensing derived classifications |
| BI | Bosque inducido | Broadleaf Forest | 41 | Not present in s1,2 |
| BJ | Bosque de tascate | Needleleaf Forest | 42 | “Juniper forest” |
| BM | Bosque mesofilo de montana | Broadleaf Forest | 41 | “Cloud forest” – technically this can contain pines, but I think it’s predominantly broadleaf |
| BP | Bosque de pino | Needleleaf Forest | 42 | “Pine forest” |
| BPQ | Bosque de pino-encino | Mixed Forest | 43 | “Pine-Oak forest” |
| BQ | Bosque de encino | Broadleaf Forest | 41 | “Oak forest” |
| BQP | Bosque de Encino-pino | Mixed Forest | 43 | “Oak-Pine forest” |
| BS | Bosque de ayarin | Needleleaf Forest | 42 | “Spruce forest” |
| BW | Bosque de bajo abierto | Broadleaf Forest | 41 | Present in s1,2 |
| DV | Sin veg. aparente | Barren Land | 31 |  |
| H2O | Cuerpo de agua | Open Water | 11 |  |
| HA | humedad | Cultivated Crops | 82 | Rainfed, Annual |
| HAP | humedad | Cultivated Crops | 82 | Rainfed, Annual, permanent |
| HAS | humedad | Cultivated Crops | 82 | Rainfed, Annual, semi-permanent |
| HP | humedad | Cultivated Crops | 82 | Rainfed, Permanent |
| HS | humedad | Cultivated Crops | 82 | Rainfed, Semi-permanent |
| HSP | humedad | Cultivated Crops | 82 | Rainfed, Semi-permanent/ permanent |
| MC | Matorral crasicuale | Shrub/Scrub | 52 |  |
| MDM | Matorral desertico microfilo | Shrub/Scrub | 52 |  |
| MDR | Matorral desertico rosetofilo | Shrub/Scrub | 52 |  |
| MET | Matorral espinoso tamauli-peco | Shrub/Scrub | 52 |  |
| MJ | Matorral de coniferas | Shrub/Scrub | 52 | Present in s1,2 |
| MK | mezquital | Shrub/Scrub | 52 |  |
| MKE | Mezquital tropical | Shrub/Scrub | 52 | Not present in s1,2 |
| MKX | Mezquital xerofilo | Shrub/Scrub | 52 | Not present in s1,2 |
| ML | chaparral | Shrub/Scrub | 52 |  |
| MRC | Matorral rosetofilo costero | Shrub/Scrub | 52 |  |
| MSC | Matorral sarcocuale | Shrub/Scrub | 52 |  |
| MSCC | Matorral sarco-crasicuale | Shrub/Scrub | 52 |  |
| MSM | Matorral sub-montano | Shrub/Scrub | 52 |  |
| MSN | Matorral sarco-crasicuale de neblina | Shrub/Scrub | 52 |  |
| MST | Matorral subtropical | Shrub/Scrub | 52 |  |
| MU | huizachal | Shrub/Scrub | 52 | Present in s1,2 |
| P/E | Pais extranjero | (nodata) | -1 | Outside the boundaries of Mexico |
| PA | Pastizal - huizachal | Grassland/Herbaceous | 71 | Present in s1,2 |
| PC | Pastizal cultivado | Pasture/Hay | 81 |  |
| PH | Pastizal halofilo | Grassland/Herbaceous | 71 | Not sure if this is anthropogenic (pasture) or just natural salt-tolerant grassland |
| PI | Pastizal inducido | Pasture/Hay | 81 |  |
| PN | Pastizal natural | Grassland/Herbaceous | 71 |  |
| PT | Veg. de Peten | Woody Wetland | 90 | Not present in s1,2 (was manglar) |
| PY | Pastizal gipsofilo | Grassland/Herbaceous | 71 |  |
| RA | riego | Cultivated Crops | 82 | Irrigated, Annual |
| RAP | riego | Cultivated Crops | 82 | Irrigated, Annual/ permanent |
| RAS | riego | Cultivated Crops | 82 | Irrigated, Annual/ semi-permanent |
| RP | riego | Cultivated Crops | 82 | Irrigated, Permanent |
| RS | riego | Cultivated Crops | 82 | Irrigated, Semi-permanent |
| RSP | riego | Cultivated Crops | 82 | Irrigated, Semi-permanent/ permanent |
| SAP | Selva alta perenni-folia | Broadleaf Forest | 41 |  |
| SAQ | Selva alta sub-perenni-folia | Broadleaf Forest | 41 |  |
| SBC | Selva baja caducifolia | Broadleaf Forest | 41 |  |
| SBK | Selva baja espinosa | Broadleaf Forest | 41 |  |
| SBP | Selva baja perreni-folia | Broadleaf Forest | 41 |  |
| SBQ | Selva baja sub-perreni-folia | Broadleaf Forest | 41 |  |
| SBQP | Selva baja sub-perreni-folia | Broadleaf Forest | 41 |  |
| SBS | Selva baja subcaduci-folia | Broadleaf Forest | 41 |  |
| SG | Selva de galeria | Broadleaf Forest | 41 | Technically wetland, but seems to have lots of trees and bushes so probably would be classified as forest in remote-sensing derived classifications |
| SMC | Selva mediana caduci-folia | Broadleaf Forest | 41 |  |
| SMP | Selva mediana perenni-folia | Broadleaf Forest | 41 |  |
| SMQ | Selva mediana sub-caducifolia | Broadleaf Forest | 41 |  |
| SMS | Selva mediana subcaduci-folia | Broadleaf Forest | 41 |  |
| TA | Temporal | Cultivated Crops | 82 | Rainfed, Annual |
| TAP | Temporal | Cultivated Crops | 82 | Rainfed. Annual/ permanent |
| TAS | Temporal | Cultivated Crops | 82 | Rainfed, Annual/ permanent |
| TP | Temporal | Cultivated Crops | 82 | Rainfed, permanent |
| TS | Temporal | Cultivated Crops | 82 | Rainfed, permanent |
| TSP | Temporal | Cultivated Crops | 82 | Rainfed, permanent |
| VA | Popal | Emergent Herbaceous Wetland | 95 | Emergent veg – herbaceous (lily/water hyacinth) |
| VD | Veg. de desiertos arenosos | Barren Land | 31 | VD areas tend to be classified as “bare” in MODIS and NLCD classifications |
| VG | Veg. de galeria | Emergent Herbaceous Wetland | 90 | “Riparian vegetation” – assume that there are shrubs and trees |
| VH | Veg. halofila | Emergent Herbaceous Wetland | 95 | Salt marsh (halophyllic veg) |
| VHH | Veg. halofila hidrofila | Emergent Herbaceous Wetland | 95 | Not present in s1,2 |
| VM | Manglar | Woody Wetland | 90 | “Mangrove swamp” |
| VP | Palmar | Broadleaf Forest | 41 | “Palm”; Present in s1,2 |
| VPI | Palmar Inducido | Broadleaf Forest | 41 | “Palm”; Not present in s1,2 |
| VPN | Palmar natural | Broadleaf Forest | 41 | “Palm”; Not present in s1,2 |
| VS | Sabana | Broadleaf Forest | 41 |  |
| VSI | Sabanoide | Broadleaf Forest | 41 | Is this a savanna? Not present in s1,2 |
| VT | Tular | Emergent Herbaceous Wetland | 95 | Cattails, reeds, etc |
| VU | Veg. de las dunas costeras | Shrub/Scrub | 52 |  |
| VW | Pradera de alta montana | Grassland/Herbaceous | 71 |  |
| VY | Veg. gipsofila | Shrub/Scrub | 52 |  |
| ZU | Zona urbana | Developed, High intensity | 24 |  |

1The NLCD classes “Deciduous Forest” and “Evergreen Forest” (codes 41 and 42) were reinterpreted here as “Broadleaf Forest” and “Needleleaf Forest”. In the United States, the distinction between deciduous and evergreen is highly correlated with the distinction between broadleaf and needleleaf. However, in Mexico, both broad- and needleleaf trees can be evergreen under the right climate conditions.

**Figures**

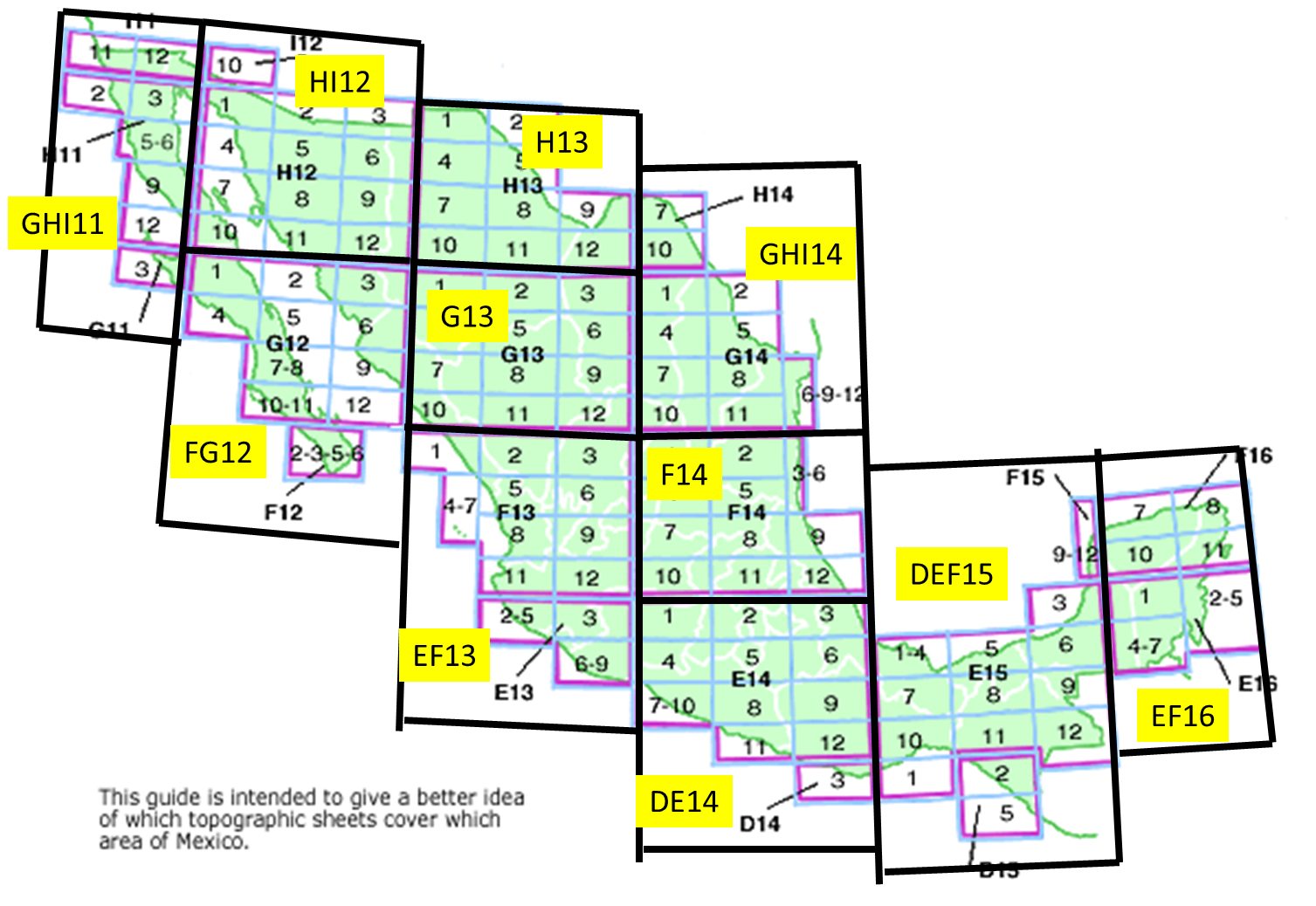


Figure 1. Tile and Meta-tile naming scheme. Meta-tile names are shown in yellow.