Analysis of Soil Survey Data in PostGIS

D. E. Beaudette

Dept. Land, Air and Water Resources
University of California
Davis, California

debeaudette@ucdavis.edu

October 27, 2008

What is PostGIS?

PostgreSQL

- Relational Database Management System (RDBMS)
- Scaleable to *n* processors, across *m* computers
- Support for very large data types and tables
- Open Source

PostGIS

- Spatial Extension to PostgreSQL
- Based on C library functions
- OGC Simple Feature Model
- Open Source (compare with $60K+/CPU for Oracle Spatial)

Why Should I Use PostGIS?

1. Scales well with massive datasets / file system objects
2. Familliar SQL-based manipulation of attribute and spatial features
3. Repeatable, transparent work-flow

SQL Review

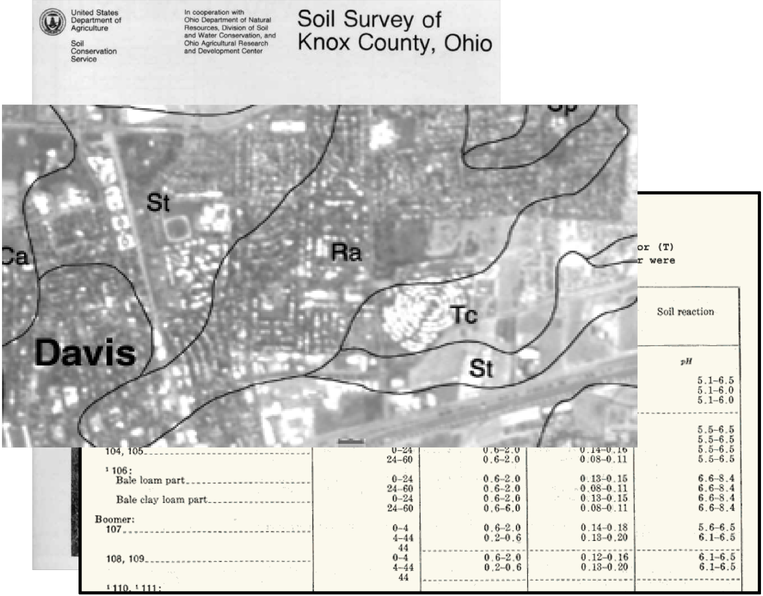
ANSI SQL Examples

- selection **select** a **from** b **where** c = d
- sorting **select** a **from** b **order by** a **desc**
- join **select** t1.a, t2.b **from** t1 **join** t2 **on** ...
- aggregation **select** **sum**(a) **from** b **group by** a

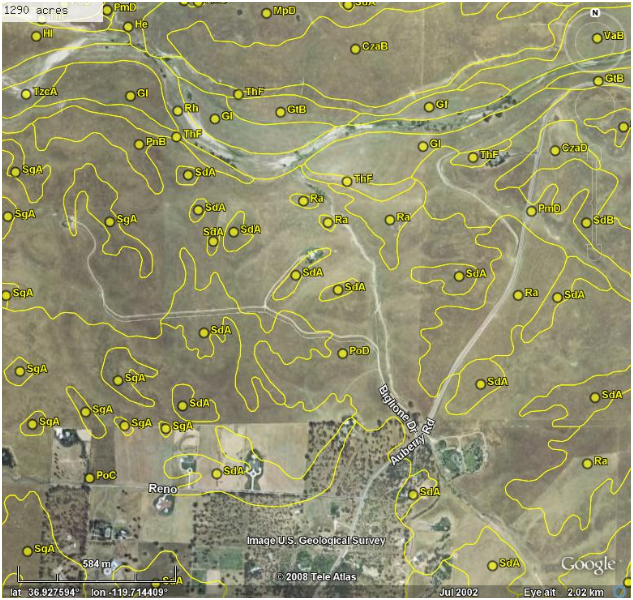
OGC "Spatial" SQL Examples

- feature extraction **select** ST_X(point_geom), ST_Y(point_geom) **from** ...
- feature extraction **select** PointN(geom) **from** ...
- spatial join **select** * **from** t1 **join** t2 **on** ST_Distance(t1.geom, t2.geom) < 100 ...
- feature manipulation **select** ST_Transform(geom, SRID) **from** ...
- feature analysis **select** ST_Buffer(geom, distance) **from** ...
- GIS overlay functions **select** ST_Intersection (geom_1, geom_2) **from** ...

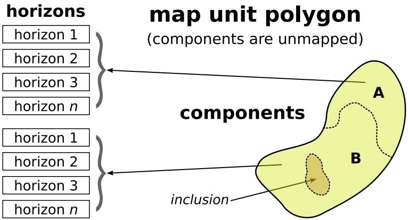
Traditional (Paper) Soil Survey

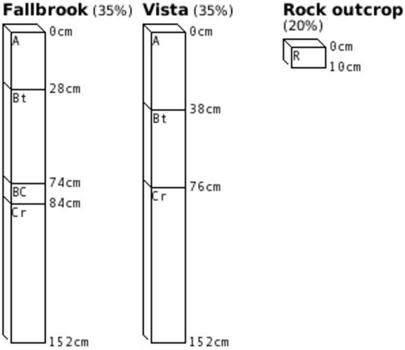


Digital Soil Survey: SSURGO

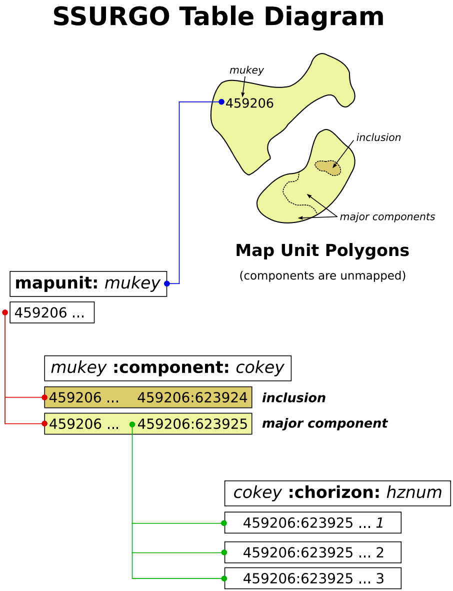


(Simplified) SSURGO Data Model (cont.)

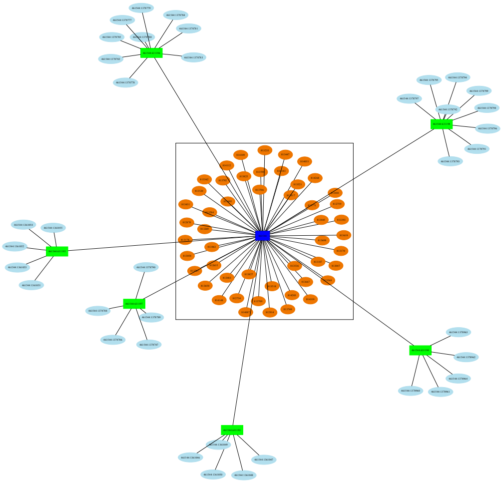




(Simplified) SSURGO Data Model (cont.)



(Simplified) SSURGO Data Model (cont.)



orange = map unit polygons, dark blue = single map unit key,
green = component keys, light blue = horizon keys

SSURGO Misc. Notes

FYI

- Any given survey must comply with basic standards, but older surveys reflect a more generalized approach than more modern surveys.
- Polygons represent a repeating pattern of legend entries (map units)
- There is a many:1:many:many (polygon:mapunit:component:horizon) relationship between spatial and horizon-level soil property data.

Metadata and Documentation (read it!)

Column Label: AWC - Representative Value (awc_r)

The amount of water that an increment of soil depth, inclusive of fragments, can store that is available to plants. AWC is expressed as a volume fraction, and is commonly estimated as the difference between the water contents at 1/10 or 1/3 bar (field capacity) and 15 bars (permanent wilting point) tension and adjusted for salinity, and fragments.[a]

[a] Soil Survey Staff. Soil Survey Geographic (SSURGO) Database for Survey Area, State. Available URL: "http://soildatamart.nrcs.usda.gov"

SSURGO Aggregation Notes

- Aggregate horizon data by one of the following methods:
  - top 1m
  - top horizon
  - profile sum
  - depth weighted (mean, median, sd)
  - most limiting
- Aggregate component by one of the following methods:
  - component percent weighted (mean, median, sd)
  - largest component (beware ties)
  - major component flag (beware ties)
  - dominant condition
- Join map unit aggregate to map unit polygons

Note

Ask a soil scientist before selecting horizon/component aggregations

Further Deatails

http://casoilresource.lawr.ucdavis.edu/drupal/node/335

Soil Profile (Horizon) Data

Query: compute the AWC for each horizon of a single component

```
select cokey, hzname, hzdept_r as top, hzdepb_r as bottom, awc_r,
(hzdepb_r - hzdept_r) * awc_r as hz_awc_cm
from chorizon
WHERE areasymbol = 'ca113'
ORDER BY cokey, hzdept_r ASC;
```

Result

cokey	hzname	top	bottom	awc_r	hz_awc_cm
757748:631223	Ap	0	15	0.18	2.7
757748:631223	A	15	31	0.19	3.04
757748:631223	Ab1	31	51	0.19	3.8
757748:631223	Ab2	51	84	0.16	5.28
757748:631223	Ab3	84	104	0.16	3.2
757748:631223	C	104	153	0.16	7.84

[...]

Component Aggregation

Query: compute the sum of horizon AWC for each component

```
SELECT cokey,
sum( (hzdepb_r - hzdept_r) * awc_r) AS component_awc_cm
FROM chorizon
WHERE areasymbol = 'ca113'
GROUP BY cokey;
```

Result

```
      cokey        | component_awc_cm
-------------------+-------------------
 459227:624019     |              6.82
 459254:624133     |             27.49
 459240:624071     |
 695115:1065805    |               3.3
 459226:624013     |              6.82
 765509:622791     |               5.4

[...]
```

Map Unit Aggregation (Step 1)

Query: summarize component-level aggregated data

```
SELECT mukey, compname, comppct_r, a.*
FROM component
JOIN
        (
        SELECT cokey,
        sum( (hzdepb_r - hzdept_r) * awc_r) AS component_awc_cm
        FROM chorizon
        WHERE areasymbol = 'ca113'
        GROUP BY cokey
        ) AS a
ON component.cokey = a.cokey
ORDER BY mukey ;
```

Result

mukey	compname	comppct_r	cokey	component_awc_cm
459275	Sehorn	85	459275:624227	14.55
459276	Sehorn	85	459276:624236	14.15
459277	Sehorn	85	459277:624237	14.15
459278	Balcom	30	459278:624243	17.86
459278	Sehorn	60	459278:624242	15.3
459279	Balcom	40	459279:624249	17.86
459279	Sehorn	50	459279:624248	14.55

[...]

Map Unit Aggregation (Step 2)

Query: compute component percentage-weighted, map unit aggregate data

```
SELECT mapunit.muname, mapunit.mukey, b.mapunit_awc_cm
FROM
        (
        SELECT component.mukey,
        sum(comppct_r * a.component_awc_cm) / sum(comppct_r) as mapunit_awc_cm
        FROM component
        JOIN
                (
                SELECT cokey,
                sum( (hzdepb_r - hzdept_r) * awc_r) AS component_awc_cm
                FROM chorizon
                WHERE areasymbol = 'ca113'
                GROUP BY cokey
                ) AS a
        ON component.cokey = a.cokey
        GROUP BY component.mukey
        ) as b
JOIN mapunit
ON b.mukey = mapunit.mukey;
```

Result

```
                    muname                          | mukey  | mapunit_awc_cm
----------------------------------------------------+--------+----------------
 Montara clay loam, 5 to 30 percent slopes          | 765511 |            5.4
 Zamora loam                                        | 459310 |             26
 Omni silty clay                                    | 459252 |          21.73
[...]
```

Aggregation of SSURGO Geometry

Query: extract SSURGO geom. from arbitrary bbox, compute area weights

```
-- define a transformed bounding box for later use
\set bbox ST_Transform( ST_SetSRID( ST_MakeBox2D(
        ST_MakePoint(-122.25033, 39.28139), ST_MakePoint(-122.20224, 39.30979)
), 4326), 9001)

-- select map unit keys, map unit symbols, and computed areas for the intersecting polygons
SELECT mukey,
sum(ST_Area(ST_Intersection(wkb_geometry, :bbox))) / ST_Area(:bbox) as mu_area_wt
from mapunit_poly
WHERE ST_Intersects(wkb_geometry, :bbox)
GROUP BY mukey;
```

Result

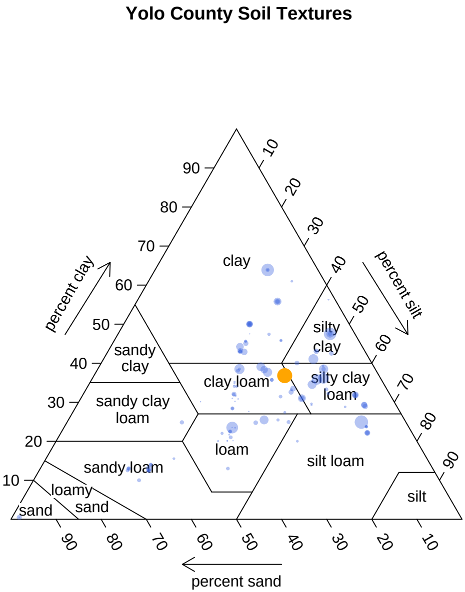
```
 mukey   |      mu_area_wt
---------+---------------------
 461544  |   0.562595368617999
 461571  |   0.347993963186697
 461595  | 0.0748614412770969
 461667  | 0.0145492269180839
(4 rows)

Time: 57.957 ms
```

Soil Texture Example

Query: compute several weighted means of soil texture parameters

```sql
-- join with polygons, and compute areas weights
SELECT mapunit_poly.mukey,
sum(ST_Area(wkb_geometry)) / (SELECT ST_Area(wkb_geometry) FROM mapunit_bound_poly WHERE areasymbol = 'ca113')
sand, silt, clay
FROM
mapunit_poly
JOIN
        (
        -- compute component percent weighted mean
        SELECT mukey,
        sum(comppct_r * sand) / sum(comppct_r) AS sand,
        sum(comppct_r * silt) / sum(comppct_r) AS silt,
        sum(comppct_r * clay) / sum(comppct_r) AS clay
        FROM
        component
        JOIN
                (
                -- compute hz thickness weighted mean
                SELECT cokey,
                sum((hzdepb_r - hzdept_r) * sandtotal_r) / sum(hzdepb_r - hzdept_r) AS sand,
                sum((hzdepb_r - hzdept_r) * silttotal_r) / sum(hzdepb_r - hzdept_r) AS silt,
                sum((hzdepb_r - hzdept_r) * claytotal_r) / sum(hzdepb_r - hzdept_r) AS clay
                FROM chorizon
                WHERE sandtotal_r IS NOT NULL
                AND silttotal_r IS NOT NULL
                AND claytotal_r IS NOT NULL
                AND areasymbol = 'ca113'
                GROUP BY cokey
                ) AS co_agg
        ON component.cokey = co_agg.cokey
        GROUP BY component.mukey
        ) AS mu_agg
ON mapunit_poly.mukey = mu_agg.mukey
GROUP BY mapunit_poly.mukey, sand, silt, clay;
```

Soil Texture Example (cont.)



Shrink-Swell (LEP) Example

Query: compute an hz-thickness weighted avg LEP for the top 1m of soil

```
-- set a lower boundary for the query
\SET lwr_bdy 100
SELECT mapunit.musym, mapunit.muname, mapunit.muacres,
round(mu_wt_lep::numeric, 2) AS lep,
CASE WHEN mu_wt_lep < 3 THEN 'Low'
WHEN mu_wt_lep >= 3 AND mu_wt_lep < 6 THEN 'Moderate'
WHEN mu_wt_lep >= 6 AND mu_wt_lep < 9 THEN 'High'
WHEN mu_wt_lep >= 9 THEN 'Very High'
END AS lep_class
FROM (
        -- compute map unit lep, weighted by component percent, to set depth
        SELECT component.mukey,
        sum(component.comppct_r * co_wt_mean_lep) / sum(component.comppct_r) AS mu_wt_lep
        FROM (
                -- compute a horizon-thickness weighted mean lep to a set depth
                SELECT cokey, sum(thick * lep_r) / sum(thick) AS co_wt_mean_lep
                FROM (
                        -- compute horizon thickness, but only to a set depth
                        SELECT cokey, hzdept_r, hzdepb_r, lep_r,
                        CASE WHEN hzdepb_r > :lwr_bdy THEN (:lwr_bdy - hzdept_r)
                        ELSE (hzdepb_r - hzdept_r) END AS thick
                        FROM chorizon
                        WHERE areasymbol = 'ca113'
                        AND lep_r IS NOT NULL
                        AND hzdept_r <= :lwr_bdy
                        ) AS hz_lep
                GROUP BY cokey
                ) AS co_lep
        JOIN component ON co_lep.cokey = component.cokey
        GROUP BY mukey
        ) AS mu_lep
JOIN mapunit ON mu_lep.mukey = mapunit.mukey
ORDER BY muacres DESC, lep DESC;
```

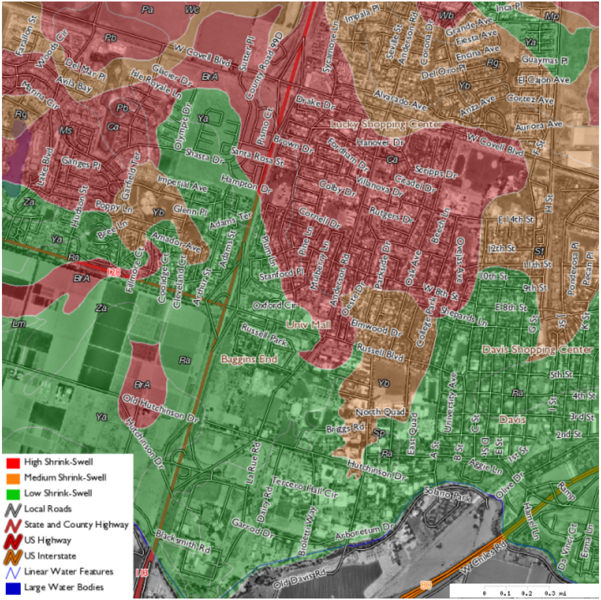
Shrink-Swell (LEP) Example (cont.)

Result: LEP for top 1meter of soil, weighed by horizon thickness and comp pct.

musym	muname	muacres	lep	lep_class
Ya	Yolo silt loam	39698	2.52	Low
Sc	Sacramento clay	34886	7.50	High
Ca	Capay silty clay	33465	7.50	High
MrG2	Millsholm rocky loam, 15 to 75 percent slopes, eroded	30118	1.50	Low
Rg	Rincon silty clay loam	24580	6.36	High
BrA	Brentwood silty clay loam, 0 to 2 percent slopes	23045	7.50	High
CtD2	Corning gravelly loam, 2 to 15 percent slopes, eroded	22080	5.34	Moderate
Mf	Marvin silty clay loam	20970	6.60	High
DaF2	Dibble clay loam, 30 to 50 percent slopes, eroded	18612	7.11	High
SmE2	Sehorn—Balcom complex, 15 to 30 percent slopes, eroded	17794	6.17	High
TaA	Tehama loam, 0 to 2 percent slopes	16622	3.75	Moderate
BdF2	Balcom—Dibble complex, 30 to 50 percent slopes, eroded	16405	5.73	Moderate
SmD	Sehorn—Balcom complex, 2 to 15 percent slopes	16117	6.50	High
BaF2	Balcom silty clay loam, 30 to 50 percent slopes, eroded	12637	4.50	Moderate
Sg	Sacramento soils, flooded	12258	6.27	High
Cn	Clear Lake soils, flooded	11666	6.92	High
SmF2	Sehorn—Balcom complex, 30 to 50 percent slopes, eroded	11226	6.33	High
Cc	Capay soils, flooded	11030	7.50	High
Sv	Sycamore complex, drained	9241	4.18	Moderate
Ms	Myers clay	8938	7.50	High
PfF2	Positas gravelly loam, 30 to 50 percent slopes, eroded	7920	5.34	Moderate
St	Sycamore silty clay loam, drained	7839	4.50	Moderate
Ck	Clear Lake clay	6946	7.50	High
Ra	Reiff very fine sandy loam	6847	1.50	Low

[...]

Shrink-Swell (LEP) Example (cont.)



Extraction of Dated Surfaces

Table: Soil series and assocaited dated alluvial formations (from Smith and Verill[1]).

Soil Series	Associated Formation	Approximate Age (1000 yrs ago)
Redding	Laguna	1600 - 2000 kya
Corning	Laguna	1600 - 2000 kya
Keyes	Laguna	1600 - 2000 kya
Whitney	Turlock Lake	500 - 700 kya
Montpellier	Turlock Lake	500 - 700 kya
Rocklin	Turlock Lake	500 - 700 kya
Snelling	Riverbank	100 - 300 kya
San Joaquin	Riverbank	100 - 300 kya
Exiter	Riverbank	100 - 300 kya
Madera	Riverbank	100 - 300 kya
Hanford	Modesto	10 - 40 kya
Grangeville	Holocene	$<$ 10 kya

[1] Smith, D.W. & Verrill, W.L. Witham, C.; Bauder, E.; Belk, D.; Ferren Jr., W. & Ornduff, R. (ed.) Vernal Pool-Soil-Landform Relationships in the Central Valley, California 1998, 15-23

Extraction of Dated Surfaces (cont.)

Make a look-up table

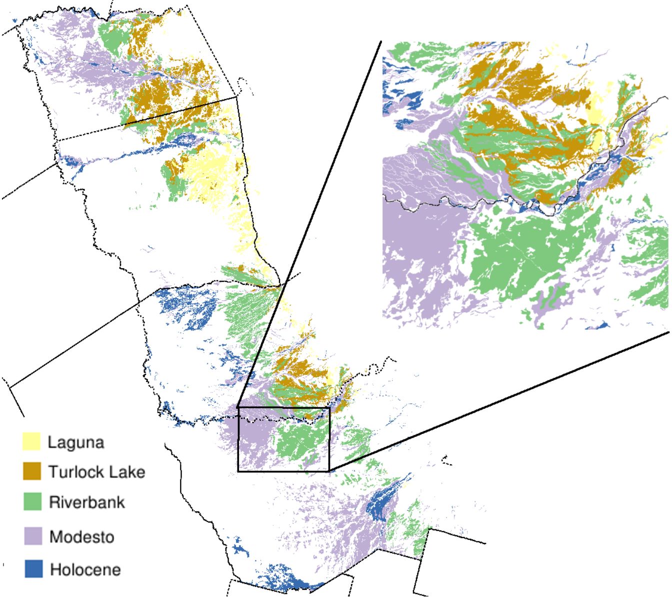
```sql
-- create a lookup table
CREATE TABLE dated_landforms (
soil_series varchar(20),
formation varchar(20),
approx_age varchar(30)
);
-- populate table
INSERT INTO dated_landforms VALUES ('Redding','Laguna','1600 - 2000 kya') ;
INSERT INTO dated_landforms VALUES ('Corning','Laguna','1600 - 2000 kya') ;
INSERT INTO dated_landforms VALUES ('Keyes','Laguna','1600 - 2000 kya') ;
INSERT INTO dated_landforms VALUES ('Whitney','Turlock Lake','500 - 700 kya') ;
INSERT INTO dated_landforms VALUES ('Montpellier','Turlock Lake','500 - 700 kya') ;
INSERT INTO dated_landforms VALUES ('Rocklin','Turlock Lake','500 - 700 kya') ;
INSERT INTO dated_landforms VALUES ('Snelling','Riverbank','100 - 300 kya') ;
INSERT INTO dated_landforms VALUES ('San Joaquin','Riverbank','100 - 300 kya') ;
INSERT INTO dated_landforms VALUES ('Exiter','Riverbank','100 - 300 kya') ;
INSERT INTO dated_landforms VALUES ('Madera','Riverbank','100 - 300 kya') ;
INSERT INTO dated_landforms VALUES ('Hanford','Modesto','10 - 40 kya') ;
INSERT INTO dated_landforms VALUES ('Grangeville','Holocene','< 10 kya') ;
```

Extraction of Dated Surfaces (cont.)

Classify soil map units based on our lookup table, return only matching map unit polygons

```
CREATE TABLE east_side_all AS
-- select geom column and the feature id
-- along with our formation names
SELECT wkb_geometry AS wkb_geometry, ogc_fid, a.*
-- the geometry table
FROM mapunit_poly
JOIN
        (
        -- list of unique mukey values joined to matching formation
        -- via soil series name
        -- DISTINCT operator is only applied to 'mukey' column
        SELECT DISTINCT ON (mukey) mukey, formation, sum(comppct_r) AS formation_pct
        FROM component
        JOIN
        -- our look-up table
        dated_landforms
        -- join condition: based on fuzzy pattern matching
        ON compname ~~* lower(soil_series || '%') = 't'
        -- additional filtering: restrict query area
        AND component.areasymbol IN ('ca654', 'ca651', 'ca649', 'ca644', 'ca648', 'ca077')
        AND majcompflag = 'Yes'
        GROUP BY mukey, formation
        -- ordering by component percent in descending order + DISTINCT = keep largest
        ORDER BY mukey, formation_pct DESC
        ) AS a
-- join condition for combining polygons with new attributes
ON mapunit_poly.mukey = a.mukey
-- limit polygon selection by survey area
AND mapunit_poly.areasymbol IN ('ca654', 'ca651', 'ca649', 'ca644', 'ca648', 'ca077');
```

Extraction of Dated Surfaces



Concluding Remarks

For Next Time

- expanded demonstration of PostGIS spatial functions
- GIS operations in the database

Additional Resources

- http://en.wikipedia.org/wiki/PostgreSQL
- http://postgis.refractions.net/documentation/
- http://casoilresource.lawr.ucdavis.edu/drupal/node/264