# Introduction

In the beginning, a process was created to do what we refer to as “aggregation”. Aggregation is the process by which a set of component attribute values is reduced to a single value to represent the map unit as a whole. This first version of the “aggregation engine” was used to create the contents of table “muaggatt” in the SSURGO data model.

Sometime later, in order to accommodate additional capabilities, a new aggregation engine was created. This version of the aggregation engine was used to perform aggregation for Windows Client Soil Data Viewer and Web Soil Survey. The original aggregation engine was still used to create the contents of table “muaggatt” in the SSURGO data model.

For some time now we have wanted to merge the older and newer aggregation engines. We finally got around to doing that, and as part of that work, we made some relatively minor changes to the capabilities of the new consolidated aggregation engine. This document discusses those changes and the impacts of those changes.

# Aggregation Engine Related Changes

1. Aggregation method “Weighted Average” can now be used for non-class soil interpretations. For example, for a map unit with the following components:

|  |  |  |
| --- | --- | --- |
| **Component** | **Representative Value Percent Composition** | **Rating for Dwelling without Basements** |
| 1 | 20 | 0.9 |
| 2 | 40 | 0.7 |
| 3 | 10 | 0.3 |
| 4 | 30 | 0.4 |

.61 would be returned (.61 = (.2 \* .9) + (.4 \* .7) + (.1 \*.3) + (.3 \* .4)).

2. Aggregation method “All Components” was renamed to “Minimum or Maximum” and its use is now restricted to numeric attributes or attributes with a corresponding domain that is logically ordered. Up to now, this aggregation method could also be selected for string attributes or attributes with a corresponding domain that was not logically ordered.

3. Aggregation method “Absence/Presence” was replaced with a more generalized version thereof, which is referred to as “Percent Present”. Up to now, aggregation method “Absence/Presence” was supported for one and only one attribute, component.hydricrating. Percent Present is a powerful new aggregation method that opens up a lot of new possibilities, e.g. “bedrock within two feel of the surface”.

4. The merged aggregation engine now supports two different kinds of horizon aggregation, “weighted average” and “weighted sum”. For the vast majority of horizon level attributes, “weighted average” is used. At the current time, the only case where “weighted sum” is used is for Available Water Capacity, where the water holding capacity needs to be summed rather than averaged.

5. The aggregation process now always returns two values, rather than one, the original aggregated result AND the percent of the map unit that shares that rating. For example, for the drainage class/dominant condition example below, the rating would be “Moderately well drained” and the corresponding map unit percent would be 60:

|  |  |  |
| --- | --- | --- |
| **Component** | **Representative Value Percent Composition** | **Drainage Class** |
| 1 | 40 | Well drained |
| 2 | 30 | Moderately well drained |
| 3 | 30 | Moderately well drained |

6. A horizon or layer where the attribute being aggregated is null will now never contribute to the final aggregated result.

7. Column sdvattribute.fetchallcompsflag is no longer needed. The new aggregation engine was updated to know that it needs to include all components whenever no component percent cutoff is specified and the aggregation method is “Least Limiting” or “Most Limiting” or “Minimum or Maximum”.

8. For aggregation methods “Least Limiting” and “Most Limiting”, the rating will be set to “Unknown” if any component has a null rating, and no component has a fully conclusive rating (0 or 1), depending on the type of rule (limitation or suitability) and the corresponding aggregation method.

9. Restrictive layers are no longer taken into account during horizon aggregation.

# Tentatively Deferred Changes

For an SDV rule, it is possible to specify both a primary and a secondary constraint column. These constraints limit what data is considered during the aggregation process. For example, for our existing crop yield SDV rules, the end user is required to select one and only one crop/yield units combination as input. For these SDV rules, it doesn’t make sense to allow the user to select more than one crop at a time.

There are cases where it does make sense to allow an end user to select more than one constraint value. For example, for column muaggatt.brockdepmin (minimum depth to bedrock), three different bedrock type constraints are specified, lithic bedrock, paralithic bedrock and densic bedrock.

The interface in both Windows Client Soil Data Viewer and Web Soil Survey wasn’t designed to allow the end user to specify more than one constraint at runtime. These interfaces could be changed to allow the selection of multiple constraints at runtime, but this a bigger change than any of the others listed above. So the recommendation of those of us here in Fort Collins would be to defer this change for now. This tentative decision is included so that our sponsors can make a final call on this issue.

The new aggregation engine is capable of dealing with multiple constraint values, and the new aggregation engine will be used to derive muaggatt.brockdepmin. This is because the XML data structure ultimately used as input to the aggregation engine supports multiple constraint values, even though the runtime user interface does not. Here is what that XML data structure looks like for muaggatt.brockdepmin:

<Column Name="brockdepmin">

<SDVRule Name="Depth to a Selected Soil Restrictive Layer" />

<PrimaryConstraint>

<Value>Lithic bedrock</Value>

<Value>Paralithic bedrock</Value>

<Value>Densic bedrock</Value>

</PrimaryConstraint>

<AggregationMethod Value="minimum or maximum" />

<ComponentPercentCutoff Value="15" />

<TiebreakRule Value="low" />

<ResultDataType Value="Int32" />

</Column>

In the work planning meeting on 4/11/2012, the decision was made that this change could be deferred to a later release.

# Implementation

## Introduction

The new merged aggregation engine will not be used until Web Soil Survey 3.0 is released. At that time a new version of NASIS and the Staging Server will also be released.

## Impact on the Process that Populates Table “muaggatt”

This process is driven by a set of SDV Rule specifications in XML format. Gary has already created this XML. A few minor tweaks, such as specifying horizon aggregation method, are still needed.

## Impact on Web Soil Survey and Windows Client Soil Data Viewer

The logic that determines the allowable aggregation methods that a user can select at runtime must be updated as follows.

If sdvattribute.algorithmname = “Percent Present” Then

Return (“Percent Present”)

ElseIf sdvattribute.attributetype = “Interpretation” and (sdvattribute.ruledesign = “Limitation” or sdvattribute.ruledesign = “Suitability”) Then

Comment: Attribute is a non-class soil interpretation.

Return (“Dominant Condition”, “Dominant Component”, “Least Limiting”, “Most Limiting”, “Weighted Average”)

ElseIf sdvattribute. mapunitlevelattribflag = 1 Then

Comment: Attribute is a map unit level attribute.

Return (“No Aggregation Necessary”)

ElseIf sdvattribute.attributelogicaldatatype = “Integer” or sdvattribute.attributelogicaldatatype = “Float” Then

Comment: Attribute is a numeric component level attribute.

Return (“Dominant Condition”, “Dominant Component”, “Minimum or Maximum”, “Weighted Average”)

ElseIf sdvattribute.tiebreakdomainname is not null

Comment: Attribute is a non-numeric component level attribute with a corresponding ordered or ranked domain. Note that this means that aggregation method Minimum or Maximum will no longer be an allowable aggregation method for character strings.

Return (“Dominant Condition”, “Dominant Component”, “Minimum or Maximum)

Else

Comment: Attribute is a non-numeric component level attribute with no corresponding ordered or ranked domain.

Return (“Dominant Condition”, “Dominant Component”)

EndIf

## Impact on the Soil Data Mart Database and the Soil Data Viewer Rule Manager Application

1. The domain associated with sdvattribute.algorithmname must be updated as follows:

1. The choice and choice label associated with choice ID 6 must be changed from “AC” and “All Components” to “MOM” and “Minimum or Maximum”, respectively.

2. The choice and choice label associated with choice ID 7 must be changed from “AP” and “Absence/Presence” to “PP” and “Percent Present”, respectively.

3, The same changes need to be made to table “sdvalgorithm” in the Soil Data Mart database, and in addition, the descriptions of aggregation methods “Minimum or Maximum” and “Percent Present” need to be updated.

These changes require changes to the NASIS Project Metadata Repository and changes to tables sdvalgorithm and sdvattribute in the Soil Data Mart database.

2. Column sdvattribute.horzaggmeth, with the corresponding domain, must be added:

|  |  |  |
| --- | --- | --- |
| Choice ID | Choice | Choice Label |
| 1 | weighted average | Weighted Average |
| 2 | weighted sum | Weighted Sum |

This requires a change to the NASIS Project Metadata Repository and a change to table sdvattribute in the Soil Data Mart database.

Note that the changes above were all addressed by release 3.0.4 of the SDM data model.

3. In the SDV Rule Manager application, when sdvattribute.horzlevelattribflag = 1, display a horizon aggregation method choice list on the Process Options tab. The default value should be “Weighted Average”. In the SDM database, sdvattribute.horzaggmeth has a default value of 1 (Weighted Average). This new control, which should be labeled “Horizon Aggregation Method”, should appear just under the control labeled “Default Aggregation Method”.

4. In the SDV Rule Manager application, whenever the user has specified a primary constraint column AND/OR a SQL WHERE Clause on the Constraints/Qualifiers tab, “Percent Present” should be added to the default aggregation choice list on the Processing Options tab.

5. A user should not be able to save an SDV Rule where the default aggregation method is “Percent Present” unless a primary constraint column AND/OR a SQL WHERE Clause has been specified on the Constraints/Qualifiers tab. When this error occurs, the following error message should be displayed:

The default aggregation method cannot be “Percent Present” when no primary constraint column or SQL WHERE Clause has been specified.

Gary points out that having both table “sdvalgorithm” and domain “algorithm” is redundant, which is now true since I have finally synced them up. I don’t propose to address this at this time, primarily to keep the number of changes to a minimum.

# Impact on a SSURGO Template Database

1, Columns “sdvattribute.fetchallcompsflag” and “SYSTEM – sdvattribute.fetchallcompflag” should be dropped.

2. Columns “sdvattribute.horzaggmeth” and “SYSTEM – sdvattribute.horzaggmeth”, both Number:Integer, should be added.

3. Import specification “sdvattribute Import Specification” need to be updated to reflect the changes to table “SYSTEM – sdvattribute”.

4. The following queries need to be updated to reflect the changes to tables “sdvattribute” and “SYSTEM – sdvattribute”:

SDV - sdvattribute - Append

SDV - sdvattribute - New Records

SDV - sdvattribute - Updated Records

SDV - sdvattribute – Update