1818 Library ◆ Suite 600 ◆ Reston, VA 20190-5602 ◆ (703) 326-2861  
<http://www.metsci.com>



SIM Land Processing   
V1.1

JUNE 4, 2013

Thomas M. Kratzke

Sr. Analyst

Contents

[1. Introduction 1](#_Toc358101910)

[2. Selection of Land Resolution 2](#_Toc358101911)

[3. Data Problems 3](#_Toc358101912)

[4. Creating the Land-Jar 4](#_Toc358101913)

[4.1 Phase‑1; Creating GSHHS++ 4](#_Toc358101914)

[4.1.1 GSHHS 4](#_Toc358101915)

[4.1.2 USGS National Atlas, NHD from ASA 4](#_Toc358101916)

[4.1.3 Running Phase‑1 4](#_Toc358101917)

[4.1.3.1 Main Class and Arguments 5](#_Toc358101918)

[4.2 Phase‑2: Refinements 5](#_Toc358101919)

[4.2.1 Providing new Refinements 5](#_Toc358101920)

[4.2.2 Running Phase‑2 5](#_Toc358101921)

[4.2.2.1 Main Class and Arguments 6](#_Toc358101922)

[5. Creating a Jar File: The entire Process 7](#_Toc358101923)

[6. Appendix: Refinements Directory 8](#_Toc358101924)

[7. References 10](#_Toc358101925)

List of Algorithms

**No table of figures entries found.**

List of Code

**No table of figures entries found.**

List of Figures

**No table of figures entries found.**

List of Formulas

**No table of figures entries found.**

List of Snippets

**No table of figures entries found.**

List of Tables

**No table of figures entries found.**

Document Revision History

|  |  |
| --- | --- |
| Version Date | Comments |
| May 23, 2013 | For *Sim* V1.5 |
| May 24, 2013 | For *Sim* V1.5 |
| June 4, 2013 | For *Sim* V1.5 |
|  |  |
|  |  |

# Introduction

This article presents a rough outline of how the land file that SIM uses is put together from various land files. SIM requires that the land data be a collection of non-intersecting directed polygons and, for each polygon, as one follows the direction of a polygon, land is to the left.

This document discusses how the files of these polygons are created, how the resolution for an individual case is selected, and how additional refinements to the land data can be made.

This article, as well as the increase of *r*1 and the introduction of the *q* resolution (see Section 2 below) was written as a response to Task #569. Task #569 requested an investigation to see if we should always use the highest resolution file. Performance tests have shown that performance is acceptable when we set *r*1to 600 *nmi*, even in the areas with the most complex coastlines. Therefore, we are using the highest resolution for the vast majority of cases, while still providing a safety net for extremely large AOIs.

In addition to producing the data files that *SIM* uses, we have also delivered a zip file of the “Working Directory” for the creation of these files. Inside of the Working Directory are shape files of all of the input data except for GSHHS, and the full resolution GSHHS input file. There are also intermediate results stored as shp files, and log files of the parts of Phase‑1 and Phase‑2 that create the intermediate results.

# Selection of Land Resolution

In subsequent sections, we will discuss how different resolutions are created. For now, suppose there are 4 different resolutions, and we’ll call them *q*, *h*, *i*, and *l*. *q* has more detail than *h*, which in turn, has more than *i*, and *l* has the least detail.

In previous versions of SIM, there was an *f* resolution which was finer than *q*. With the increased refinement in the Chesapeake Bay, New York Harbor, and the Great Lakes, we smoothed that resolution ever so slightly to create the *q* resolution.

SIM uses a jar file containing the land data. This jar file must be in SIM’s classpath, and we will call it the “land-jar.” For each resolution, there are 2 files in the land-jar, and their names correspond to that resolution.

When the land-jar is created, we decide which resolutions we wish to include. SIM will consider only those resolutions in the land-jar. Currently, we are using the above 4 resolutions.

To determine which resolution is used for each case, SIM uses two parameters in Sim.properties; call them *r*1 and *r*2, where *r*1<*r*2. For a given case, SIM finds the smallest circle surrounding the AOI. Let *r* be its radius. If there are two available resolutions in the land-jar, then SIM chooses the more detailed if *r*<(*r*1+*r*2)/2. If there are three resolutions, SIM chooses the finest resolution if *r*<*r*1, the coarsest if *r*>*r*2, and the middle resolution otherwise. If there are 4 or more resolutions, SIM splits the middle section between *r*1 and *r*2, by using a logarithmic scale.

In SIM 1.5, we have increased *r*1 from 240 *nmi* to 600 *nmi*, and kept *r*2 at 3000 *nmi*. In 1.4, we were using the *f* resolution and, in the worst case, this could result in noticeably longer processing times. In 1.5, we built the *q* resolution by using our program that builds lower resolutions. The *q* resolution is barely less fine than the *f* resolution, and it reduces the extra-fine detail given in NY Harbor, Chesapeake Bay, and the Great Lakes.

# Data Problems

We start with a few preliminary statements about data. The algorithms that combine different sources of data require non-intersecting (and hence nested) polygons. No source of data has satisfied this requirement, including GSHHS. Hence, we have written and applied a program that identifies problems and “cleans up” the data.

The most difficult problem to address is a polygon that crosses itself or another. A self-crossing polygon is the more difficult of these two to deal with, because it is unclear which section of the polygon is intended as the real data. We log the situation and report the problem. Until the problem is resolved, we simply take the larger of the two polygons, and discard the rest. Thus far, this has been adequate.

Other problems that are cleaned up in this pre-processing include duplicated points, “spikes,” and polygons that have edges that coincide for a distance. We identify those, log them, and then use our own algorithms to eliminate those.

# Creating the Land-Jar

We use two programs to prepare the files for the land-jar, and we call them Phase‑1 and Phase‑2. Phase‑1 takes the current GSHHS data (2.2.2 as of May, 2013) and merges USGS and NHD data. These are relatively stable sets of data. Phase‑2 takes the result of Phase‑1 and “refinements” from ASA to produce a new “full-resolution” GSHHS-like file. Then Phase‑2 also creates lower-resolution GSHHS-like files.

## Phase‑1; Creating GSHHS++

### GSHHS

Our main source of land data is the GSHHS files [Wessel and Smith, 1996]). GSHHS stands for “Global Self-consistent Hierarchical High-resolution Shorelines.” It is a very concise collection of the world's shorelines stored as *complete polygons*, which is very handy.

GSHHS also forms the basis of the coastline support in the Generic Mapping Tools (GMT), which is used by over 10,000 users worldwide (gmt.soest.hawaii.edu; Wessel and Smith [1998]). It is supported by the National Science Foundation, and is released under the GNU General Public License. (See <http://www.gnu.org/copyleft/gpl.html>).

### USGS National Atlas, NHD from ASA

In addition to the GSHHS data, we use two and possibly three other sources of data in Phase‑1. These datasets provide better inland water data than GSHHS. GSHHS, although it has some lakes and reservoirs, is primarily intended for the world’s shorelines, so we supplement the GSHHS data with another source of data. We use this data to process lakes, streams, canals, and reservoirs for the United States and its territories.

The first set of data is from the United States Geological Survey (USGS). The USGS data is refined by using data called “Region 17.” This data is supplied by ASA. Region 17 data is primarily around the Columbia River. The USGS data is modified as per the Region 17 data, and this data is combined with the GSHHS data.

If additional islands are provided in the form of kml or shp files, they can be incorporated at this step. This feature is currently not being used. There were two islands south of Long Island in New York that were provided, but those have been built into the most recent versions of GSHHS.

Note that although GSHHS comes in different resolutions, we deal only with their highest (or “full”) resolution file.

### Running Phase‑1

Phase‑1 creates a GSHHS-like data file called hydroGp\_gshhs-f, using the data mentioned so far. For debugging purposes, we can create lower resolution versions of hydroGp\_gshhs-f, but for production purposes, there is no need.

Using a 2011 desktop, Phase‑1 takes about 2 hours to run. Phase‑1 produces results in the Working Directory’s subdirectories “Phase1LogDir,” “Phase1Results,” and “Phase1ShpFiles.”

#### Main Class and Arguments

Phase‑1 is run by running the java app whose main class is com.metsci.util.gshhs.buildAll.Phase1BuildAll. The arguments are:

1. rawData/Region17\_Area\_Dissolve/Region17\_Area\_Dissolve.shp
2. 0.05
3. rawData/Usgs/hydrogp020.shp
4. rawData/gshhs/gshhs\_f.b

The first argument specifies the Region17 shp file and the second is a parameter that specifies that we don’t want the full resolution of the Region17 data. The third argument specifies the shp file of the USGS data, and the fourth argument specifies the GSHHS file that we use. Note that the last parameter is not a shp file. Note also that all of these are input specifications. The output specifications (where the results go) are simply hardcoded.

## Phase‑2: Refinements

ASA provides “refinement files” for the program Phase‑2. For each section of the coast that we wish to “refine,” we actually completely “redefine it.” ASA provides two shp files for each section; a “mask” file and a “detail” file.

The mask is used to define which region is to be altered. In particular, any point that is not within any polygon of the mask, is considered to be on land or in water, as per the result coming out of Phase‑1.

In practice, we take all of the polygons from both the detail and mask and treat this as the mask. The first step of Phase‑2 is to change everything in the mask to water. We then fill in land (and water, in the case of lakes within land polygons) by using the polygons from the detail file.

### Providing new Refinements

To provide a new region of refinement, ASA provides a Mask shp file and a Detail shp file. All Mask/Detail shp file pairs are processed in one pass. Hence, the mask will consist of all polygons that are in *any* Mask shp file, and the set of Detail polygons will be all of the polygons in the Detail shp files. Hence, new refinements are treated just as old refinements are; they are treated together in one large update process.

The directory structure for the Refinements directory is given in the appendix.

### Running Phase‑2

Phase‑2 is the program that takes the file hydroGp\_gshhs-f that Phase‑1 created, and produces a full-resolution version that incorporates the refinements. The result is a “full-resolution” file, but we do not use that within SIM. Phase‑2 produces lower resolution files from the full-resolution file. Currently, we are creating the *q*, *h*, *i*, and *l* resolutions, and those are the ones that are eventually included in the land-Jar.

Phase‑2 runs in about 9 hours. If a very low resolution is asked for, the time can run to 16 hours. Low resolutions take much more time than high resolutions because more points are dropped and there are more opportunities for crossings to occur.

#### Main Class and Arguments

Phase‑2 is run by running the java app whose main class is com.metsci.util.gshhs.buildAll.Phase2BuildAll. The arguments are:

1. The name of the hydroGp\_gshhs full resolution file
2. The directory that contains the Mask/Detail directory pairs
3. The resolutions.

A resolution is specified by a number and a letter. The number gives the coarseness of the resolution, and the letter is the name. The higher the number, the coarser is the resolution. For example, we are currently using: “0.01 q 0.05 h 0.25 i 1.25 l” for our resolutions. The number 0.01 is associated with the *q* resolution, 1.25 is associated with the *l* resolution, and the rest follow this pattern.

The current list of arguments is:

1. Phase1Results/hydroGp\_gshhs\_f.b
2. rawData/Refinements
3. 0.010 q 0.05 h 0.25 i 1.25 l

In the appendix, we list the entire directory structure of the Refinements subdirectory. This was obtained by using the dos command “dir /s” from within Refinements.

# Creating a Jar File: The entire Process

We have little control over GSHHS and so we simply run Phase‑1, logging any strange polygons. Problems that occur when cleaning up the raw GSHHS file, are reported to GSHHS, but we simply work with our own cleaned up version. Currently, there are spikes in several polygons, and Paul Wessel of GSHHS is reviewing them.

We then run Phase‑2, and here we do have some control. Phase‑2 logs anomalies and we report these to ASA. ASA then responds to the anomalies by producing a cleaned up version of the refinements. We continue working with ASA with this until we have no more anomalies to report. At that point, we will use the program Phase‑2 to create the lower resolutions.

To create the land-jar file, we jar the result of Phase‑2. These include the full resolution and lower resolution data sets. For each resolution, Phase‑2 produces a pair of files, one with a suffix of ‘b’ and one with a suffix of ‘d.’ We put all of these b-files and d-files into the directory <SomeDirectory>/com/metsci/util/gshhs/data, and make sure that there is nothing else in <SomeDirectory>. Then we put <SomeDirectory> into our java path and jar up <SomeDirectory>, naming it whatever we wish to name it. That jar file can then go into SIM’s java path, but we must delete any other land-jar file. When done, SIM is ready to use the new land-jar.

# Appendix: Refinements Directory

05/11/2013 12:31 PM <DIR> Chesapeake\_NY

05/11/2013 12:31 PM <DIR> GreatLakes

05/11/2013 12:31 PM <DIR> puget\_sound

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\Chesapeake\_NY

05/11/2013 01:08 PM <DIR> Detail

05/11/2013 08:54 AM <DIR> Mask

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\Chesapeake\_NY\Detail

05/11/2013 05:49 AM 28,006 Land\_Ches\_NY.dbf

05/11/2013 05:49 AM 145 Land\_Ches\_NY.prj

05/11/2013 01:08 PM 38,868 Land\_Ches\_NY.qix

05/11/2013 05:49 AM 27,484 Land\_Ches\_NY.sbn

05/11/2013 05:49 AM 1,156 Land\_Ches\_NY.sbx

05/11/2013 05:49 AM 13,760,808 Land\_Ches\_NY.shp

05/11/2013 05:49 AM 7,109 Land\_Ches\_NY.shp.xml

05/11/2013 05:49 AM 22,452 Land\_Ches\_NY.shx

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\Chesapeake\_NY\Mask

05/11/2013 05:49 AM 127 Water\_Ches\_NY.dbf

05/11/2013 05:49 AM 145 Water\_Ches\_NY.prj

05/11/2013 08:54 AM 64 Water\_Ches\_NY.qix

05/11/2013 05:49 AM 132 Water\_Ches\_NY.sbn

05/11/2013 05:49 AM 116 Water\_Ches\_NY.sbx

05/11/2013 05:49 AM 13,567,568 Water\_Ches\_NY.shp

05/11/2013 05:49 AM 9,094 Water\_Ches\_NY.shp.xml

05/11/2013 05:49 AM 108 Water\_Ches\_NY.shx

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\GreatLakes

05/11/2013 01:47 PM <DIR> Detail

05/11/2013 09:38 AM <DIR> Mask

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\GreatLakes\Detail

05/11/2013 05:50 AM 428,207 GL\_LAND\_Simple.dbf

05/11/2013 05:50 AM 145 GL\_LAND\_Simple.prj

05/11/2013 01:47 PM 49,132 GL\_LAND\_Simple.qix

05/11/2013 05:50 AM 32,740 GL\_LAND\_Simple.sbn

05/11/2013 05:50 AM 1,060 GL\_LAND\_Simple.sbx

05/11/2013 05:50 AM 8,652,620 GL\_LAND\_Simple.shp

05/11/2013 05:50 AM 10,709 GL\_LAND\_Simple.shp.xml

05/11/2013 05:50 AM 27,932 GL\_LAND\_Simple.shx

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\GreatLakes\Mask

05/11/2013 05:50 AM 454 GL\_Water\_Simple.dbf

05/11/2013 05:50 AM 145 GL\_Water\_Simple.prj

05/11/2013 09:38 AM 64 GL\_Water\_Simple.qix

05/11/2013 05:50 AM 132 GL\_Water\_Simple.sbn

05/11/2013 05:50 AM 116 GL\_Water\_Simple.sbx

05/11/2013 05:50 AM 8,468,376 GL\_Water\_Simple.shp

05/11/2013 05:50 AM 8,019 GL\_Water\_Simple.shp.xml

05/11/2013 05:50 AM 108 GL\_Water\_Simple.shx

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\puget\_sound

05/11/2013 01:55 PM <DIR> Detail

05/11/2013 09:46 AM <DIR> Mask

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\puget\_sound\Detail

05/11/2013 05:50 AM 43,358 puget\_sound\_land.dbf

05/11/2013 05:50 AM 145 puget\_sound\_land.prj

05/11/2013 01:55 PM 9,408 puget\_sound\_land.qix

05/11/2013 05:50 AM 6,668 puget\_sound\_land.sbn

05/11/2013 05:50 AM 420 puget\_sound\_land.sbx

05/11/2013 05:50 AM 3,061,444 puget\_sound\_land.shp

05/11/2013 05:49 AM 7,260 puget\_sound\_land.shp.xml

05/11/2013 05:49 AM 5,332 puget\_sound\_land.shx

D:\kratzke\kratzke\Geo\dev\WorkingDirectory2\rawData\Refinements\puget\_sound\Mask

05/11/2013 05:49 AM 270 puget\_sound\_water.dbf

05/11/2013 05:49 AM 145 puget\_sound\_water.prj

05/11/2013 09:46 AM 64 puget\_sound\_water.qix

05/11/2013 05:49 AM 132 puget\_sound\_water.sbn

05/11/2013 05:49 AM 116 puget\_sound\_water.sbx

05/11/2013 05:49 AM 3,070,824 puget\_sound\_water.shp

05/11/2013 05:49 AM 1,656 puget\_sound\_water.shp.xml

05/11/2013 05:49 AM 108 puget\_sound\_water.shx

# References

Wessel, P., and W. H. F. Smith (1996), A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res., 101(B4), 8741–8743.

Wessel, P., and W. H. F. Smith (1998), New, improved version of Generic Mapping Tools released, Eos Trans., AGU, 79(47), 579.