

OPPORTUNITIES FOR GENERALIZATION
IN THE
DIGITAL CHART OF THE WORLD

by

Frank James Fico

NCGIA Technical Report 92-2

August 1992

Table of Contents

| | |
|--|----------|
| List of Figures | page iii |
| Abstract | page v |
| Acknowledgements | page vi |
| Introduction | page 1 |
| Literature Review | page 3 |
| DCW Generalization Opportunities | page 9 |
| A. Political/Oceans | page 10 |
| B. Drainage | page 12 |
| C. Drainage-Supplemental | page 13 |
| D. Hypsography | page 14 |
| E. Hypsography-Supplemental | page 16 |
| F. Vegetation | page 16 |
| G. Physiography | page 17 |
| H. Ocean Features | page 18 |
| I. Populated Places | page 19 |
| J. Roads | page 20 |
| K. Railroads | page 22 |
| L. Transportation Structure | page 22 |
| M. Utilities | page 23 |
| N. Land Cover | page 23 |
| O. Landmarks | page 24 |
| P. General Culture | page 25 |
| Q. Aeronautical | page 25 |
| Summary | page 26 |
| Appendix A: Figures | page 29 |
| Appendix B: DCW Features in the Study Area | page 74 |
| Appendix C: Generalization Operator Descriptions | page 78 |
| References | page 81 |

List of Figures

| | |
|---|---------|
| Figure 1: unmodified Political/Oceans layer | page 30 |
| Figure 2: unmodified <i>land</i> (unfilled), simulating <i>coastline, dejure international boundaries</i> and <i>coastal closure shoreline</i> | page 31 |
| Figure 3: coarsened <i>land</i> (unfilled), simulating coarsened <i>coastline, dejure international boundaries</i> and <i>coastal closure shoreline</i> | page 32 |
| Figure 4: unmodified Drainage layer | page 33 |
| Figure 5: unmodified <i>streams/rivers</i> | page 34 |
| Figure 6: <i>streams/rivers</i> remaining after omission | page 35 |
| Figure 7: unmodified <i>perennial inland water</i> | page 36 |
| Figure 8: <i>perennial inland water</i> remaining after omission | page 37 |
| Figure 9: unmodified Drainage-Supplemental layer | page 38 |
| Figure 10: unmodified Hypsography layer | page 39 |
| Figure 11: unmodified <i>1000 to 3000 feet</i> and <i>3000 to 7000 feet</i> | page 40 |
| Figure 12: <i>1000 - 3000 feet</i> and <i>3000 -7000 feet</i> after WEEDDRAW operation | page 41 |
| Figure 13: unmodified Hypsography-Supplemental layer | page 42 |
| Figure 14: unmodified Vegetation layer | page 43 |
| Figure 15: unmodified <i>vegetation</i> | page 44 |
| Figure 16: <i>vegetation</i> after WEEDDRAW operation | page 45 |
| Figure 17: unmodified Physiography | page 46 |
| Figure 18: <i>escarpments, bluffs, cliffs, etc.</i> found along the coast | page 47 |
| Figure 19: "escarpmented coastline" | page 48 |
| Figure 20: unmodified Ocean Features layer | page 49 |
| Figure 21: unmodified <i>rocks, isolated or awash</i> | page 50 |
| Figure 22: "rocky coastline" | page 51 |
| Figure 23: unmodified <i>coral reefs</i> | page 52 |

| | |
|--|---------|
| Figure 24: "reefed coastline" | page 53 |
| Figure 25: unmodified Populated Places layer | page 54 |
| Figure 26: unmodified <i>built-up areas</i> | page 55 |
| Figure 27: <i>built-up areas</i> after WEEDDRAW operation | page 56 |
| Figure 28: unmodified total point features in Populated Places layer | page 57 |
| Figure 29: unmodified <i>populated places (names in city tints)</i> | page 58 |
| Figure 30: unmodified Roads layer | page 59 |
| Figure 31: unmodified <i>primary/secondary roads or highways</i> | page 60 |
| Figure 32: unmodified Railroads layer | page 61 |
| Figure 33: Railroads layer without <i>connectors</i> | page 62 |
| Figure 34: unmodified Transportation Structure layer | page 63 |
| Figure 35: unmodified Utilities layer | page 64 |
| Figure 36: <i>power transmission lines</i> remaining after omission | page 65 |
| Figure 37: unmodified Land Cover layer | page 66 |
| Figure 38: unmodified <i>undifferentiated wetlands</i> | page 67 |
| Figure 39: <i>undifferentiated wetlands</i> after WEEDDRAW operation | page 68 |
| Figure 40: unmodified <i>unconsolidated materials</i> | page 69 |
| Figure 41: <i>unconsolidated materials</i> after WEEDDRAW operation | page 70 |
| Figure 42: unmodified Landmarks layer | page 71 |
| Figure 43: unmodified General Culture layer | page 72 |
| Figure 44: unmodified Aeronautical layer | page 73 |

Abstract

The Digital Chart of the World (DCW) is a database built from 1:1,000,000 scale source maps by the Defense Mapping Agency (DMA), scheduled for public release in early 1992. Nearly 200 feature types are included in 17 data layers. A layer-by-layer examination of features included in a prototype area of the DCW is accomplished for the purpose of investigating the feasibility of building a smaller scale database via generalization. Opportunities for feature generalization are identified and specific generalization operations are described. A number of the operations are performed or simulated using ARC/INFO, and the results are graphically displayed. It is found that meaningful small scale feature representation can be achieved through the application of tailored generalization operations with a significant savings in the amount of data required.

KEY WORDS: Digital Chart of the World (DCW), generalization

Acknowledgements

This monograph reports on Master's thesis research completed at NCGIA-Buffalo, during academic year 1990-1991. The thesis was supervised by Barbara P. Battenfield, and forms a portion of NCGIA Research Initiative 8, "Formalizing Cartographic Knowledge". Funding to support Mr. Fico's graduate work was provided by the U.S Defense Mapping Agency, as part of the DMA Long-Term Instruction Program. Permission to publish this thesis has been granted by DMA Public Affairs Officer David L. Black (15 October, 1991). The views expressed in this paper are those of the author and do not necessarily represent positions of the Defense Mapping Agency or the U.S. Department of Defense.

Introduction

The Digital Chart of the World (DCW) is a new product of the Defense Mapping Agency (DMA) scheduled for public release in early 1992. It is a digital database of nearly 200 feature types populated from the 1:1,000,000 scale Operational Navigation Chart (ONC) series, DMA's largest scale product for which worldwide coverage is available (except Antarctica). The features consist of vector points, lines and areas, and thus can be selected for individual display by feature type. Features are grouped by theme in 17 "layers," not including a separate layer of metadata (data lineage) information. The DCW database, consisting of 10-15 gigabytes of information, will be available on four CD-ROMs. Although DCW is not a geographic information system (GIS) per se, it is designed to be a database compatible with many commercial GISs. Simple display operations can be performed without a GIS link on IBM-compatible personal computers using software written by Environmental Systems Research Institute (ESRI), DMA's prime contractor on the DCW project.

The primary purpose of this project, a multinational effort involving agencies of Australia, Canada and the United Kingdom along with DMA, is the introduction of "a set of vector product standards oriented toward the Geographic Information System environment" (DMA, 1990). It is hoped that future digital products will adhere to this so-called vector product format (VPF), facilitating data exchange. Already, DMA has produced VPF prototypes of other digital products that can be displayed using the same ESRI software.

Release of DCW is being anxiously awaited by numerous users of GIS and digital data. A list of potential applications for DCW includes a "backdrop for national databases, global and theater plans and assessments, briefing and decision graphics, index for spational [*sic*] retrieval of other data, and global physical modeling" (DMA, 1990).

Undoubtedly, additional applications have been envisioned by many organizations, both in the public and private sector.

DCW actually consists of two databases: the "detailed database" is derived from the ONC product (1:1,000,000 scale source) and is normally referred to as "DCW", while the "browse database" (1:15,000,000 scale) exists to allow the user to navigate within the detailed database. DCW users may be expected to require some subset of the detailed data to be viewed in a regional or continental area of interest, roughly from 1:4,000,000 to 1:15,000,000. As it exists, DCW has no capability to provide a coherent visual presentation of most of its included features at this scale of interest. When features collected at 1:1,000,000 are displayed at smaller scales, crowding and coalescing inevitably occurs. And while the browse database provides limited base map features appropriate to a world view, the features become very sparse with much intermediate blank space when zoomed in to a larger scale.

My research relates to the creation of a third database within DCW to allow for a complete transition from a worldwide view down to scale of 1:1,000,000; in effect, filling the "data gap" that presently exists within the intermediate scale range. Rather than building a new database from scratch, I propose a subset of the detailed database be generalized to populate the "gap fill" database at an arbitrary target scale of 1:4,000,000. I will outline the layers and features appropriate for inclusion and describe the generalization operations required to display the features at one quarter the compilation scale with maximum visual acuity. To the extent possible, I will also perform the recommended operations on a copy of prototype 4 of the DCW that has been converted to ARC/INFO format and show the "before" and "after" depictions of each feature.

Literature Review

Generalization has been the subject of numerous papers in the discipline of cartography. Recent literature has addressed the subject with regard to issues in GIS and digital databases. Muller (1990b) outlined the underlying "motivations" for generalization in today's environment in terms of various requirements. The economic requirement controls the amount of data populating the original ungeneralized database, and has direct bearing upon the extent of generalization necessary or desirable. Data robustness is addressed through generalization as data errors are smoothed out and basic trends emerge. Graphics generated by modern GIS and spatial decision support systems often are utilized for decision-making, and make up the display and communications requirements for generalization. Finally, modern digital databases like DCW are generally designed for multiple applications at a variety of scales, and generalization is necessary to fulfill these multipurpose requirements.

The answer to "why do we generalize?" in the context of a digital environment was attempted by McMaster and Shea (1988). Their categorization of philosophical, application and computational objectives provides a checklist of guidelines or goals to achieve with generalization. Philosophical objectives include reducing complexity, maintaining both spatial and attribute accuracy, maintaining a logical hierarchy, consistently applying generalization rules and maintaining aesthetic quality. The last objective seems rather nebulous. The selection of the appropriate scale given the map purpose and intended audience, and the maintenance of clarity make up the application objectives. Their computational objectives stress maximizing the performance of algorithms and amount of information retained while minimizing data storage and the algorithms' memory requirements.

A follow-on paper describes mapping situations when conditions warrant generalization (Shea and McMaster, 1989). The authors define six conditions that may

occur with scale reduction: congestion, coalescence, conflict, complication, inconsistency and imperceptibility. Because these conditions are often highly subjective and difficult to quantify, certain measures are available to aid in the determination of the conditions. Measures such as distance, length and density are quite distinct, but categories of distribution, shape, gestalt measures and abstract measures are seemingly just as subjective and vague as some of the conditions.

Many researchers have sought to refine the broad concept of generalization into manageable pieces. Perhaps the aspect of generalization most relevant to this paper is digital generalization, defined as "the application of both spatial and attribute transformations in order to maintain clarity, with appropriate content, at a given scale, for a chosen map purpose and intended audience" (McMaster and Shea, 1988, p. 242). With digital databases emerging as the backbone of current cartographic and GIS applications, the formalizing of digital generalization methods will most likely occupy much of the literature in the near future. Automated map design systems will rely upon a formalized knowledge base to perform generalizations upon digital databases when necessary (Beard and Mackness, 1991). The challenge is to specify and implement a digital version of the decision rules, often highly ambiguous and inadequate in the world of manual cartographic generalization (Zoraster *et al.*, 1984).

Other distinctions have been made within the realm of digital generalization. Brassel and Weibel (1988) consider "cartographic generalization" to encompass only those operations affecting the displayed version of the database; operating upon the stored information was called "statistical generalization." Alternatively, statistical generalization could be thought of as primarily emphasizing the positional accuracy while cartographic generalization worked on "visual effectiveness or recognizability" (Buttenfield, 1989, p. 79-80). Mark (1989) replaces "cartographic" with "graphical" and further refines this category: "visual generalization" is composed of general procedures that operate equivalently on both geographic and non-geographic features, while

"geographical generalization" takes the characteristic geometric structure of the feature into account before generalizing. Some workers have simply refined the concept based on the nature of the generalization operators. For example, Muller (1990b) divides operators into "geometric" versus "conceptual" generalization, and Weibel (1987) uses "filtering" to mean generalization by elimination or simplification .

At the heart of the generalization process is the selection and application of the proper generalization operator. The literature abounds with operator descriptions (e.g. Battenfield and Mark, 1990; Weibel, 1987; Shea and McMaster, 1989; Beard and Mackaness, 1991). By no means is there a consensus opinion regarding a standardized list of operators, whose numbers range from three to 12. This paper will utilize the eight operators of Beard and Mackaness (1991): select, omit, coarsen, collapse, combine, classify, exaggerate, displace. Because the definitions of these operators may vary from one cartographer to another, see Appendix C for a complete description of how each is used here.

The notion that generalization should be applied using criteria specific to the geographic phenomena present is prevalent in the recent literature. The dominant rule in Mark's (1989) geographic generalization is that preservation of geographic relations among features is paramount during generalization, at the expense of positional accuracy when necessary. A familiar example of this is the case of a stream, road and railroad all passing through a narrow canyon. When represented on a small scale map, the three features might be collinear. Performing a geographic generalization requires the crossing and adjacency relations among the three features be retained, most likely necessitating exaggeration of the width of the canyon and some displacement of the linear features. The case of a winding mountain road undergoing a coarsening operation provides another example. A standard algorithm might output an artificially straightened road that fails to convey the geographic character of the situation. A true geographic generalization would forego the standard coarsening and represent the road with a sinusoidal symbol,

sacrificing positional accuracy but successfully imparting the geographic setting to the user.

Before any phenomenon-based generalization operation can be successful, properties specific to certain geographic features must be identified. Buttenfield (1989) describes cartographic lines as either self-similar (whose depictions are unchanging with scale; e.g. roads) or scale dependent (e.g. rivers). Knowing which category a particular linear feature belongs to can aid the selection of appropriate generalization operators and tolerance values. Buttenfield asserts that cartographic lines could be distinguished by this property. As an example, she uses Richardson plots to separate coastlines into erosional (self-similar) and depositional (scale dependent) types.

Once an operator has been chosen, digital generalization is performed through a particular algorithm. Because the majority of cartographic information is represented as lines, most published algorithms work on linear features. A good summary of line generalization algorithms can be found in Zoraster et. al. (1984), a report to review and recommend algorithms suitable for DMA's cartographic production. Algorithms are placed into nine classes based on type, and the report finds algorithms employing tolerance bands, point relaxation and digital elevation model (DEM) smoothing to be of greatest potential utility to DMA. There are also a few examples in the literature of algorithms to generalize objects other than lines. Muller (1990a) presents algorithms to remove point-point and point-line collisions resulting from line generalization. A solution for point clustering, also applicable to text placement and line labeling, is found in Mackaness and Fisher (1987).

Generalization algorithms usually require a tolerance value specifying the relative degree of generalization to accomplish. Consider the following algorithms:

"if point2 is separated from point1 by $< x$, combine point1
and point2"

"if line1 has length $< x$, omit line1"

"if linewidth $< x$, exaggerate linewidth"

In each of the preceding examples, "x" is the algorithm's tolerance value. Careful thought must be given in choosing the tolerance value that provides the appropriate reduction of detail. An aid to determining tolerance values is a line's "structure signature" (Buttenfield, 1991), which consists of five parameters designed to uniquely characterize a linear feature. The structure signature is a measure of variation in line geometry, and can help fine-tune selection of tolerance values for features displaying multiple geometrical trends.

An examination of feature geometry over a systematic scale change is performed by Muller (1990b). At map (or display) scales of less than 1:10,000, isomorphic representation of features is possible; that is, all objects can be faithfully drawn to scale. Some form of generalization is necessary at smaller scales. Geometric generalization, encompassing the operations of simplification, enlargement and displacement, dominates from 1:20,000 to 1:200,000, transitioning to the conceptual generalization operations of selection, classification, typification and symbolization above 1:500,000. Within the two transition zones (1:10,000-1:20,000 and 1:200,000-1:500,000), Muller describes the representation of a particular object experiencing a "catastrophic change," where a small scale variation causes a large variation in the object geometry. The first catastrophe occurs when a previously isomorphic representation is forced to undergo some degree of geometric generalization; for example, road widths are exaggerated and buildings are simplified, combined and displaced. The movement from geometric to conceptual generalization causes the second catastrophe, marked by the symbolization of a town by a circle, or the selection of a subset of the total drainage network. The exact point of catastrophe is thought to be feature dependent, but deserves further research.

The idea of "catastrophic change" is highly relevant to the issue of multiple representations. The fact that different linear features require highly variable amounts of data points to store at a given scale presents a dilemma for digital storage: either a single

database is stored at a given scale, incorporating redundant information for some features and insufficient information for others, or multiple databases of varying scales are stored, increasing the amount of data and creating a data management problem. Researchers are pursuing the idea of a single, "scaleless" database (Muller, 1990b; Bittenfield and DeLotto, 1989) in which only topology need be stored, but its attainment is many years away. Knowledge of scales of catastrophic change for various features can aid all three of these storage solutions.

Mark (in Bittenfield and DeLotto, 1989, p.69) advocates a single database including all possible detail, putting the onus on the user to "zoom in" past the inevitable clutter and indistinguishable blobs of data to an area of interest. Not only is this often cumbersome for the user, but the large amounts of data involved fill valuable disk space and slow response times. Mark admits that multiple databases will predominate as the near-term solution to the multiple representation problem. What follows is a description of efforts to convert DCW to a multiple database product. Generalization of the detailed database may serve the needs of future DCW users for data views at multiple scales.

DCW Generalization Opportunities

The following discussion provides a layer-by-layer examination of what presently populates the DCW database at a 1:1,000,000 scale and what types of generalization operations might be performed to create a new database appropriate to a 1:4,000,000 scale of representation. The layers are ordered by their relative importance to a typical user of the smaller scale database. Regardless of application, coastlines and international boundaries are mandatory base map features to establish a visual orientation; hence, the political/oceans layer is addressed first. The drainage and hypsography layers are covered next to extend an overall geomorphology to the base map. Vegetation, physiography and ocean features are layers to embellish the landscape that has been established, and they complete the "natural" layers of the database. Next, the cultural layers are added, starting with populated places and the transportation networks and infrastructure connecting them (roads, railroads and utilities layers). The land cover layer, while including some natural features, is predominantly composed of agricultural and extraction-based feature types, and therefore appears following the initial cultural layers. Finally, the narrow interest layers of landmarks, general culture and aeronautical are treated with the lowest priority.

Analysis was performed on a portion of DCW Prototype 4 converted to ARC/INFO format for DMA internal use only. The geographic coverage includes the United Kingdom, western France and the northern Iberian peninsula. This is the region covered by ONC's E1 and F1, two of the four charts making up Prototype 4. Appendix B includes a table showing all the DCW features included in this region. This is the pool from which features were selected for generalization. DCW layers generally correspond to ARC/INFO coverages displayed using ARCPLOT, and individual features are selectable by Boolean operation using the feature-type code. The coverages were converted to Transverse Mercator from the default Plate Carree projection, because less distortion was

apparent at the latitude of the study region. Except where noted otherwise, all the graphics were generated at a scale of 1:4,000,000 from ARCPLOT plot files converted to POSTSCRIPT format and printed in black-and-white on a laser printer. Print formatting was performed using the Superpaint and Powerpoint packages on a Macintosh. Only the simplest point, line and area-fill symbols were used to construct the maps to maximize visual clarity. Each map caption includes a number representing the amount of digital information displayed, to the nearest kilobyte.*

A. Political/Oceans (figure 1)

This is probably the single most useful layer for the vast majority of DCW applications. Specifically, the international boundaries and coastline features are the primary items of interest to any user wishing to construct an informative base map at small scale. The other features are not required for the gap-fill 1:4,000,000 database for various reasons. *Small islands*, experimentally included in ONC F1 only and thus not visible on figure 1, would be omitted as extraneous information, as would the lines outlining the prototype boundary. The *land* and *ocean* area features provide the capability to fill these areas with contrasting colors to enhance visual presentation. Since these "features" basically fill existing areas defined by the coastline features, they may be retained as "nice to have" with little additional storage overhead.

Three features remain after the preliminary selection operation to populate the database (figure 2), all of which can sustain a simpler representation through the application of a coarsening operation without sacrificing a critical loss of cartographic information. *Coastlines* form the major feature layer of the three, and the geographic area including the complex active margin of northwestern Scotland is considered to provide

*amount of information shown is for the entire 8.5 X 11 page; insertion of margin space for binding will reduce this slightly

ample opportunity to coarsen. It happens that the single international boundary located within this study area (Ireland-Northern Ireland) is defined by geography, rather than set equal to a particular latitude or longitude value. Thus, this feature also benefits from coarsening. The third feature, *coastal closure shoreline*, is essentially a connector added during the digitization process to close the coastline across river mouths. Though the coarsening operation has little effect upon this minor feature, its presence facilitates the secondary generalization operation of omission (see below).

The ARCPLLOT command WEEDDRAW was used to coarsen, with the tolerance value set to .3 (non-essential points composing the lines that are within .3 inches of another at the display scale are eliminated - see ESRI, 1989). WEEDDRAW applies the algorithm described in Douglas and Peucker (1973), a cartographically sound tolerance band type routine (McMaster, 1987). Residual to this operation are numerous small islands and narrow peninsulas that have been reduced to the point of approximating one- or zero-dimensional objects. It is convenient to be able to omit these residuals based on some minimum area threshold; however, displaying line features does not allow for area attributes. A solution is to display the *land* feature (which does have an area value) without filling the polygons. Because *land* includes country codes as an attribute, international boundaries are included to define separate *land* polygons. And by definition, *land* includes the *coastal closure shoreline* feature in order to exist as closed polygons. Hence, figure 2 is actually the 323 unfilled *land* polygons, which corresponds to the three line features of interest. Subsequent to the coarsening operation, all polygons having area less than .258 square km were omitted from the display. Figure 3, then, shows the 12 land polygons surpassing the area threshold and coarsened at the .3 tolerance level, and represents the generalized version of the political/oceans layer. Its content may be classified into just two features: "coastline" and "international boundaries." Note that the data content of figure 3 is twenty times less than that of figure 2.

B. Drainage (figure 4)

This layer suffers from extreme crowding when displayed at 1:4,000,000. Clutter within this layer can be attributed to the *streams/rivers* and *inland shorelines* line features and *perennial inland water* area feature. *Dams* are non-conflicting at this scale, and can be passed along to the new database without modification, as can *canals, aqueducts, etc.* Features eliminated by selection include *connector, none* and *inland water island*. *Inland shorelines*, used to bound inland water, would also be excised because it provides no information that cannot be displayed with the inland water features. *Non-perennial inland water* would be retained, with the assumption that generalization operations similar to those performed on *perennial inland water* may be required in arid environments. *Streams/rivers* and *perennial inland water* therefore remain in need of generalization.

Figure 5 shows all the *streams/rivers* over most of the United Kingdom. An ideal way to generalize this feature would be to omit all those streams of less than some threshold length, then collapse the double-lined stream segments (included as *perennial inland water* area feature) into *streams/rivers*. After "weeding out" the smaller streams, the overall basin pattern (more appropriate to a smaller scale) will become easier to discern. However, to accomplish this, the streams must be digitized as complete entities to include total length as an attribute. Automatic digitization employing anything short of full artificial intelligence techniques precludes the digital storage of true river length. The scanning process used for DCW creates a drainage network composed of numerous, short stream segments from which total stream length is impossible to derive. Therefore, this feature was manually generalized by omission to simulate a length threshold operation. Rather than a length threshold, streams not included in a world atlas map of approximately 1:5,000,000 scale (Rand McNally, 1968, p. 4) were omitted from the new database.

Figure 6 is a manually drawn representation of this operation, created by tracing those streams found in the atlas off of figure 5 using ARCEDIT. It is quite obvious that the streams of figure 6 have been naturally coarsened by manual digitization at a smaller scale. While this is a desirable operation to eliminate excess detail and should be performed in an automatic generalization scheme, it remains secondary to the omission operator for this feature, which is not nearly as complex as the coastline feature. Undoubtedly, both individual streams and overall drainage basins are more easily delineated with this generalized version of *streams/rivers*.

Perennial inland water (figure 7) is the other drainage feature in need of generalization. Simply by selecting an appropriate minimum area threshold, all the double-lined stream segments included in this feature can be omitted, as well as undesired lakes. However, to be consistent with the *streams/rivers* generalization performed above, only the lakes included in the atlas are retained for the gap-fill database, classified as "lakes" (figure 8). Coarsening of the polygon borders is not appropriate because unlike most other DCW area features, *perennial inland water* has very distinct edges, defining a unique geometry that should be preserved whenever possible.

C. Drainage-Supplemental (figure 9)

This is an experimental DCW layer, including features from ONC F1 only. The area covered on figure 9 extends from western France at the upper right (northeast), across the Pyrenees, and into the northern Iberian peninsula. (The area devoid of features in the northwest quadrant is the Bay of Biscay.) Most of the features displayed are *small lakes*. Some clustering of the features can be seen in northern France and in the vicinity of the Pyrenees, indicating this layer could benefit from generalization. However, this layer consists solely of drainage features included on the ONC but smaller than the DCW minimum polygon size of 3.14 mm in circumference (ESRI, 1990, p. 34). Considering

that numerous larger polygons of inland water have been omitted from representation in the gap-fill database in a previous operation, it is inappropriate to include any features from this layer. Any small lake of special significance should be included in the landmarks layer.

D. Hypsography (figure 10)

This layer contains over 1 Mb of information; too much to be able to print on the Apple LaserWriter II Postscript printer. Figure 10 is an ARC PLOT plot file sent directly to a Calcomp electrostatic plotter. Point features consist of *spot elevations*, which by their nature are collected in a very well-distributed pattern. Though there are numerous *spot elevations*, their dispersal allows these features to be replicated without modification to the gap-fill database. The line features are dominated by *closed land contours*, included at 1000-foot intervals. The lines are highly complex, displaying more convolutions in the Scotland area than the coastline there (compare to figure 2). The complexity of contours is aggravated by their mandatory closure. Often, nearby contour loops create additional interference in a small scale representation already experiencing line clutter. Perhaps due to this unique geometric character, the WEEDDRAW operation failed to satisfactorily coarsen the contours at various tolerance levels. All line features in this layer will be deleted from the new database.

Like those of the political/oceans layer, the area features of hypsography are space-filling. These features provide an elevation tint capability beyond that found in the source ONC's. The hypsography area features can be considered a generalized version of the layer's line features. Elevations ranging from 0 to 11,000 feet are represented using a contour interval of 1000 feet, providing enough information for twelve elevation zones, yet only five are included. This is a classification operation in conjunction with what might be considered a reverse collapse operation, in that information formerly represented with a one-dimensional symbol is now represented using a two-dimensional

symbol. Shaded elevation zones is the representation of choice for the hypsography layer of the new database.

Rather than displaying a single elevation zone at a time, all zones above a certain threshold elevation can be selected. For example, in figure 11, the shaded area represents all elevations greater than 1000 feet by printing the *1000 to 3000 feet* and *3000 to 7000 feet* features together. (The two highest zones are found on ONC F1 only.) The elevation zone displayed in figure 11 is much easier to interpret at this scale than the contour lines found in figure 10. However, figure 11 can benefit from additional generalization. Hypsography is being treated like a type of land cover, and therefore should borrow generalization techniques employed in the other land cover layers. Specifically, polygons separated by less than a threshold distance should be combined, polygons smaller than an area threshold should be omitted, and the boundaries of those that remain should be coarsened. None of these operations is sanctioned by the literature (Mark, 1989), which recommends generalizing the underlying DEM whenever simplified contours are sought. DCW hypsography, however, is not based upon any DEM, nor is it likely the source ONC used a DEM to construct contour lines. It is fair to assume the original contours in DCW have been generalized from the point at which they were manually compiled on an ONC, making this feature's boundaries more indistinct than other land cover type features.

The WEEDDRAW command, when used to coarsen filled polygons, simulates the other two operations described above. As excess points are removed, the coarsened polygon boundaries often incorporate nearby smaller polygons, approximating a combine operation. And when a small polygon is coarsened to the point of being unable to contain any fill pattern, it is omitted from the display. Figure 12 shows the result of this operation using a tolerance of .05 when applied to figure 11. The smallest polygons retained on the new image are about 10 square km in area. This new representation is more compatible with the generalized coastline feature (figure 3). The tolerance value

will most likely need to be adjusted when treating a different geographic area with its unique topography, or when other elevation thresholds are desired.

E. Hypsography-Supplemental (figure 13)

Figure 13 is the other graphic derived directly from a plot file due to the extreme density of information in this layer. Like the supplemental drainage layer, this layer adds hypsographic features more appropriate to a larger scale representation. The single feature, *partial intermediate or auxiliary contour*, has a much greater geographic extent and is more complex than closed land contours from the previous layer. No area features are present to facilitate a generalization scheme described above. The sheer magnitude of data displayed suggests an intensive generalization operation would be necessary to render this layer comprehensible at this scale. None will be attempted, however, because this layer will not appear among the candidate gap-fill layers and features. Certainly, the utility of this layer is restricted to display scales larger than even the DCW source scale of 1:1,000,000.

F. Vegetation (figure 14)

Although hypsography was treated as a land cover for generalization purposes, this is the first actual land cover layer considered. It is experimental as well, and found only on the ONC F1 area. Figure 14 covers the same geographic area as figure 9 (the drainage-supplemental layer). Two area features comprise this layer: *vegetation* and *hole in vegetation/none*. The latter comprises a very minor portion of the layer (compare figure 14 to figure 15, showing *vegetation* only), and is of no use to the new database. *Vegetation* only will be generalized.

The series of three operations useful for most land cover polygons was mentioned in section D: polygons separated by less than a threshold distance are combined, polygons smaller than an area threshold are omitted, and the boundaries of those that

remain are coarsened. Unfortunately, there is no easy way in ARC/INFO to perform the combine operation, which ought to precede the others. The ARC command DISSOLVE performs an aggregation operation, but the polygons are required to be mutually exclusive/collectively exhaustive in order that they have common boundaries. Given a raster data structure, an alternative could be a clustering type operation borrowed from image processing, whereby the x largest polygons are chosen, then pixels of the same value are systematically assigned to one of the seed polygons. For the situation at hand, the WEEDDRAW command is found to create a fair simulation of this series of operations (see description in section D). Surely, algorithms can be coded to perform the operations exactly as stated, but that is beyond the scope of this paper.

As was accomplished for hypsography, WEEDDRAW is enacted upon figure 15, displaying all *vegetation*. Again, the tolerance value is .05. Unsurprisingly, the resulting figure 16 captures a geometry similar to that found on the final hypsography image (figure 12). While the amount of information present in this layer has been more than halved in the generalized version, the reduction is not so dramatic as it had been for previous layers. This is probably due to the fact that 95% of the original data was retained for generalization, a much higher proportion than seen previously.

G. Physiography (figure 17)

This layer consists of two highly segmented line features. *Levees, dikes, eskers* accounts for just a single segment, and can be passed to the new database with no generalization. Most of figure 17 consists of *escarpments, bluffs, cliffs, etc.* in the southern England area. Strictly speaking, this feature does not require any generalization, since all the segments have a simple geometry and are clearly separable at this scale. And with the entire layer consisting of only 3 Kb of information, data reduction is certainly not a valid reason for generalization.

However, an opportunity exists to reduce the clutter and enhance visualization based on a special property of the cliffs feature. Figure 18 shows those cliffs found along the coastline. It appears very similar to the entire layer (figure 17). Actually, only three segments have been deleted: the one levee feature and the only two inland cliff segments. Those cliffs found along the coastal margin can be classified as "escarpmented coastline," while the few inland cliffs can be retained in the original feature class. In addition, the new feature class should undergo a linear combination operation using the rule that any coastline segment of at least 40 km in length containing over 50% *escarpments, bluffs, cliffs, etc.* shall be classified as "escarpmented coastline." The implementation of this operation produces a graphic (figure 19) better able to visually summarize the gist of the data populating this layer. Ideally, the new feature would be built by overlaying the cliffs onto the coarsened coastline produced in section A in order that the generalized coastal geometry be maintained.

H. Ocean Features (figure 20)

This name of this layer is somewhat misleading. Rather than including features throughout ocean areas, only features within a narrow coastal margin populate this layer. A more descriptive (though somewhat cumbersome) name might be "coastal margin marine features," plainly evidenced by the outline of the British Isles and northern France that can be seen in figure 20. *Exposed wrecks* are retained without modification.

The other point feature, *rocks, isolated or awash*, show no real opportunity for generalization in figure 20. However, along the northern and northeastern margins of the Iberian peninsula, rocks are concentrated enough to overlap (figure 21, displaying rocks only). Their extreme collinearity with the coastline suggests a reverse collapse generalization, using the rule specifying all coastal segments of at least 50 km in length that contain 10 or more rocks shall be classified as "rocky coastline." Rocks not meeting the generalization criteria are retained in the original feature class, but dense

concentrations form this new line feature. The two charter members of "rocky coastline" appear in figure 22.

Reefs (figure 23) can be thought of as the geomorphological inverse of coastal cliffs, thus they are generalized in an identical manner. Coastal reefs are linearly combined and classified into "reefed coastline" (figure 24) using the rule formulated for cliffs (section G). The few reef segments not found along the coast plus those coastal reefs not generalized retain their membership in *reefs*. The unique tightly crenulated line symbol used for reefs is only apparent at much larger scale, but those features retained in the original class should be coarsened to eliminate the crenulations and reduce information content. As noted for the new "escarpmented coastline" feature class, both "rocky coastline" and "reefed coastline" should also be registered to the coarsened coastline feature of the political/oceans layer.

I. Populated Places (figure 25)

As the initial coverage of DCW's cultural information, it is important to establish an appropriate cultural base map of populated places. The layer consists of both point and area features. The *none* area feature, representing empty areas within city tints, is insignificant and would not be transferred to the new database. At first glance, figure 25 appears to require little or no generalization at this scale. This scope of this report does not include treatment of the annotation found in certain DCW layers, but populated places is the layer to which place name labeling is the most important. Imagine a text label for every place on figure 25 and the need for generalization of this layer becomes clear.

The *built-up areas* feature (figure 26) represents a pool from which the generalized layer should be built. In the hierarchy of populated places, this feature logically depicts those places large enough to require an area, rather than a point symbol. An omission operation based on a population threshold would most likely treat this feature as

representative of the more populous places, omitting the point features. But *built-up areas* also represents a type of cultural land cover, in that boundary lines have been drawn to arbitrarily limit the extent of this rather abstract attribute value. Thus, it will be treated with the same generalization scheme utilized for the physical land cover polygons. The WEEDDRAW routine (with .05 tolerance) omits a multitude of small polygons, while again simulating the combine and coarsen operations as well (figure 27).

The point features are not completely eliminated from consideration. Though the *populated places* feature is not included in the new database, the second point feature is retained in its entirety. Selecting only this feature results in a subset of 267 points (figure 29) out of the 1533 total point features present in both features (figure 28). This subset of points is retained for two reasons. These points exclusively represent place names located within the bounds of *built-up areas*; in other words, where the area tint has engulfed the place point symbol. Therefore, they are necessary to define a precise location of a place name within the built-up zone. More importantly, the inclusion of these points effects a collapse operation, whereby those areas omitted in the WEEDDRAW operation are restored as point symbols. The original visual hierarchy of point and area symbols, communicating a general sense of population, is thus replicated in the generalized version of this layer.

J. Roads (figure 30)

The roads layer is the major cultural infrastructure network of interest. The road line features are hierarchical, including *dual lane (divided) highways, primary/secondary roads or highways* and *tracks, trails, footpaths* (not included in the ARC/INFO DCW version). This hierarchical arrangement of features is in contrast to that found in drainage, the major physical network addressed in section B. *Streams/rivers*, the primary feature of that layer, is an all-inclusive class containing far too many members for coherent display at this scale (see figure 5). A logical breakdown of drainage might

include streams/rivers as a layer, with first-order streams, second-order streams, etc. as feature classes within the layer to reduce the number of members found in the current mega-class *streams/rivers*. With such a feature arrangement, the roads layer facilitates a generalization by selection operation. For example, figure 31 shows *primary/secondary roads or highways* only.

The ability to select road subsets is a major reason this layer is retained without modification for the gap-fill database. Another important reason is the characteristic morphology of a cultural line feature such as roads, again contrasted to that of the primary physical line feature, streams. The most obvious difference between the two features is that man-made lines are much simpler and smoother in appearance due to inherent engineering constraints; thus, dense cultural line networks like roads are easier to visually comprehend than dense drainage lines (compare figures 5 and 30).

Beyond the physical appearance of the lines, knowledge of what the lines represent makes the road network more understandable. In viewing road lines, one must only visualize in two dimensions; roads are simply connections between points. However, drainage lines must be visualized in three dimensions because rather than simple connections, streams inherently describe flow direction. Interpreting a stream map invariably involves either conscious or subconscious drainage basin delineation, a formidable task given the complexity of figure 5. The generalization operation chosen for *streams/rivers* omitted most of the data, but the resulting features adequately conveyed the major drainage basin boundaries (figure 6). There are no "road basins" remaining after omitting a subset of roads; each segment present within *primary/secondary roads or highways* carries equal weight in describing that feature class. Each road line omitted is an important piece of information lost. Since figure 30 contains no major feature conflicts, no roads are omitted.

K. Railroads (figure 32)

Most of the properties characterizing roads also hold true for railroads. The engineering constraints with respect to maximums of radius of curvature and grade are even more restrictive in the case of railroads, meaning railroad lines should be less complex than roads. Railroads are also much fewer in number than roads because railroad corridor planning involves more variables and decisions than road planning. And like the roads layer, railroads are sub-divided into single and multiple track feature categories, providing the ability to generalize by selection. Despite being less dense than roads, railroads display prominent clustering around the transportation "hubs" of London, Birmingham and Manchester. Eliminating *connectors* nicely expunges these hubs (figure 33), though line crowding within the hubs is not a problem in figure 32. Retention of these lines better conveys the character of the railroad feature without creating a resolution conflict. Hence, this layer will transfer unmodified to the new database.

L. Transportation Structure (figure 34)

Another experimental layer digitized from ONC F1 only, this is an attempt to separate point and short line transportation structure features from the line networks found in the roads and railroads layers. The feature type categories are not specific, but a status value for each member identifies exactly what kind of structure is being represented. Figure 34, covering the same view of western France and northern Spain and Portugal as the vegetation study area, primarily contains railroad tunnels, which could use some generalization in the Barcelona area (southeast). However, no operations are attempted because, although it is topologically important to maintain this layer for larger scale applications, there is no reason to include it in a gap-fill 1:4,000,000 database.

M. Utilities (figure 35)

This final cultural network-type line coverage presents a spate of generalization problems to which there are no straightforward solutions. Perhaps because there is no connector feature in utilities, the lines are not as easy to follow as roads, which contains much more data (compare figure 35 to figure 30). The transportation networks both allowed selection to help thin the network out, but *power transmission lines*, like *streams/rivers*, offers no feature sub-categories to select. As is the case for all the line features in this prototype, automatic digitization precludes the digital identification of single, contiguous power lines, with which some kind of omission rule could be built (*e.g.*, omit all power lines carrying less than x kilowatts). When all else failed in the case of streams, an atlas map of close to the target scale was used as a source for an omit operation, but there is no similar small scale source for power lines readily available.

A small advantage found in the power line display is that there are none included within large built-up areas. This absence enables figure 35 to remain mostly legible, despite being visually "busy." Figure 36 has omitted all the disconnected segments of less than 40 km, and presents a somewhat less disorganized display of *power transmission lines*. Again, this is merely a compromise operation to keep the depiction of utilities on the order of the road and railroad networks, which were not changed. The best solution for these layers and drainage lines is to manually capture whole objects during the digitization process in order to facilitate flexible automated generalization operations farther down the line.

N. Land Cover (figure 37)

The geographic area covered by figure 37 is approximately the same portion of the British Isles as outlined in the political/oceans layer (figure 1). This layer is reserved for a number of disparate land cover types, the majority of which are agricultural- or extraction industry-based. However, within this area of the world, only a few feature

classes are present. The *mines* point feature along with the *salt pans* area feature have no need for generalization, and the *none* area feature is deleted as irrelevant to the new database. Remaining, then, in need of generalization are *unconsolidated materials* and *undifferentiated wetlands*, two area features having a distinctly more natural than cultural flavor. For this particular case, land cover would more appropriately be considered with the natural layers, but its component features would generally place it here, among the cultural layers.

Unsurprisingly, the operation applied to all previous land cover-type area features is also utilized here. The wetlands feature (figure 38) is very similar to vegetation (figure 15), and the WEEDDRAW command produces output as expected (figure 39). *Unconsolidated materials* (figure 40) has a distribution more akin to that of populated places (figure 26), though not as highly clustered. In fact, this polygon distribution is probably the most irregular and scattered of all those to which WEEDDRAW is applied. The output (figure 41) is the least satisfactory, at least in a strictly qualitative sense. Clusters of smaller polygons in figure 40 react to the operation somewhat inconsistently, producing unexpected results in some cases in figure 41. It may be that the WEEDDRAW simulation of polygon combination, omission and coarsening used throughout this study works best for relatively large, clustered polygons.

O. Landmarks (figure 42)

As the catch-all layer for landmark features of all kinds, this layer has no set of rigid, pre-defined feature codes. The highly disguised geographic area presented in figure 42 is the same sections of France, Spain and Portugal displayed for vegetation, drainage-supplemental, and transportation structure because landmarks is the last of the four experimental layers derived solely from ONC F1. With such a flexible content, this layer may present views where generalization is required, though obviously none is needed here. Assuming most areas will not have high concentrations of landmarks, this layer

will pass on unchanged to the new database. However, if there are any members of feature classes restricted from the gap-fill database (*e.g. small lakes* or *small islands*) that are unusually strategic or significant, they may be preserved within this layer.

P. General Culture (figure 43)

The many feature codes included in this layer are quite varied, ranging from *towers* to *trading posts* to *international date line*, and they must be examined individually for potential generalization operations. Line features typically are characterized by very short segments that are represented smaller than the point symbol size of .05 inches at this scale, and should not be included in the gap-fill database due to their insignificance. Other line features (*e.g. aerial cableways*) might be better represented collapsed to point features. The individual features *weirs*, *jetties*, *ramps*, *breakwaters*, *piers*, *wharfs* and *quays* could perhaps be classified into a new point or area feature type "marine structures" or "port." Most of the point features, along with *international date line*, should be retained without modification. Only one feature type is found in the prototype: *prominent fences*. As figure 43 shows, this is one of the insignificant features that is deleted from the new database.

Q. Aeronautical (figure 44)

At this time, just one feature is specified to occupy the aeronautical layer: the *airport* point feature. In keeping with DCW's stated general-purpose objective, no specialized aeronautical information is included. To be sure, associated with *airport* are considerably more attribute data than found with other features, but of course this has no bearing on the size of the symbols, which describe a nice distribution over the British Isles on figure 44. This layer is moved undisturbed to the new database.

Summary

Features present in a prototype version of the DMA Digital Chart of the World have been studied in detail to determine where generalization opportunities exist to build a third database suited to a smaller scale view. In some cases, entire coverage layers can be deleted as inappropriate to a smaller scale representation (drainage-supplemental, hypsography-supplemental, transportation structure). Conversely, other layers should undergo no generalization because their included elements are relatively few and important to retain (landmarks, aeronautical). The remainder of the layers should be generalized to a certain degree.

A preferred means of omitting elements from the four line network layers (drainage, roads, railroads, utilities) based on a size threshold is not performed due to lack of an adequate degree of topology in the digitally scanned data. However, only in the *streams/rivers* feature of drainage is omission necessary for easy visualization of the feature at small scale. In this case, the operation is simulated by manually reconstructing streams found in an atlas (assuming only these pass a particular size threshold) and omitting the remainder. Little or no generalization is attempted on the three cultural line network layers because each is very much free of visual conflicts, and are still quite usable at the target scale of 1:4,000,000. However, omission by threshold rules would be applied if the data structure allowed, simply to reduce the amount of data stored in the gap-fill database.

The Douglas-Peucker (1973) line simplification algorithm is used to coarsen coastline and international boundary features of the political/oceans layer via the WEEDDRAW command in ARCPLLOT. This same routine is utilized in a less conventional manner to simulate polygon combination, omission and edge coarsening for a variety of area features in the hypsography, vegetation, populated places and land cover layers, where it appears to perform best on large polygons that are filled and somewhat

clustered. WEEDDRAW provides a quick and easy way to generalize what are envisioned to be some of the most frequently displayed DCW features, including those defining both the natural and cultural "base map."

The omission and coarsening operations account for the generalization of most of the major features. Within the physiography and ocean features layers, the collapse and classification operations are used to create the new generalized feature types of "escarpmented coastline," "rocky coastline" and "reefed coastline." These operations occur on relatively minor features and their small scope of coverage allows them to be accomplished manually in the ARCEDIT subsystem. Although the data reduction associated with the collapse operations is relatively minor, the small scale visual communication of these features' morphology is greatly enhanced.

Indeed, along with improving visual interpretability, data reduction is an important goal of creating a third DCW database. The reduction of the amount of data required to represent the gap-fill database from that found in the detailed DCW database is substantial. While it is difficult to determine the exact amount of digital storage needed by each of the databases based solely on a study of part of a modified prototype of DCW, an estimate can be derived from the amount of data shown on each figure. The single greatest contribution to data elimination is the deletion of the hypsography-supplemental layer, which accounts for over one megabyte of data in figure 13 alone. A comparison of the "before" and "after" views of generalized features shows the WEEDDRAW coarsening operation saves a total of more than 1 Mb as well. The remaining operations contribute close to an additional Mb of data savings, of which about half are the manual omissions performed in the drainage layer. Thus, the total savings that can be verified from analysis of the figures is over 3 Mb, which does not include individual features deleted from layers undocumented by printed output. The fact that the information content of the 17 figures depicting the unmodified DCW layers totals 5.147 Mb implies

this rough rule of thumb: the gap-fill database should contain just 40% of the information found in the detailed database.

DMA and other government agencies have recently begun to expend prodigious amounts of resources building digital databases that can be utilized by geographic information systems. Attention is often devoted to providing enough information to support the largest possible scale of representation over the greatest area. To be sure, this attention is quite justified due to the many applications requiring such a detailed view. But certainly there are an equal number of applications benefiting from a smaller scale, regional view. Until the single, scaleless database concept (Muller, 1990b; Buttenfield and DeLotto, 1989) becomes an operational reality, separate databases are required to fully support these applications. Rather than building them from scratch, the large scale data that has been digitized at great expense should be generalized to create the smaller scale databases. More effort is needed dedicated to developing generalization schemes tailored to specific data requirements and target scales. This paper, as a preliminary study of the generalization opportunities available in the DCW detailed database, provides direction for further development of a third DCW database that could better address the needs of many potential users of the product upon its public release.

Appendix A: Figures

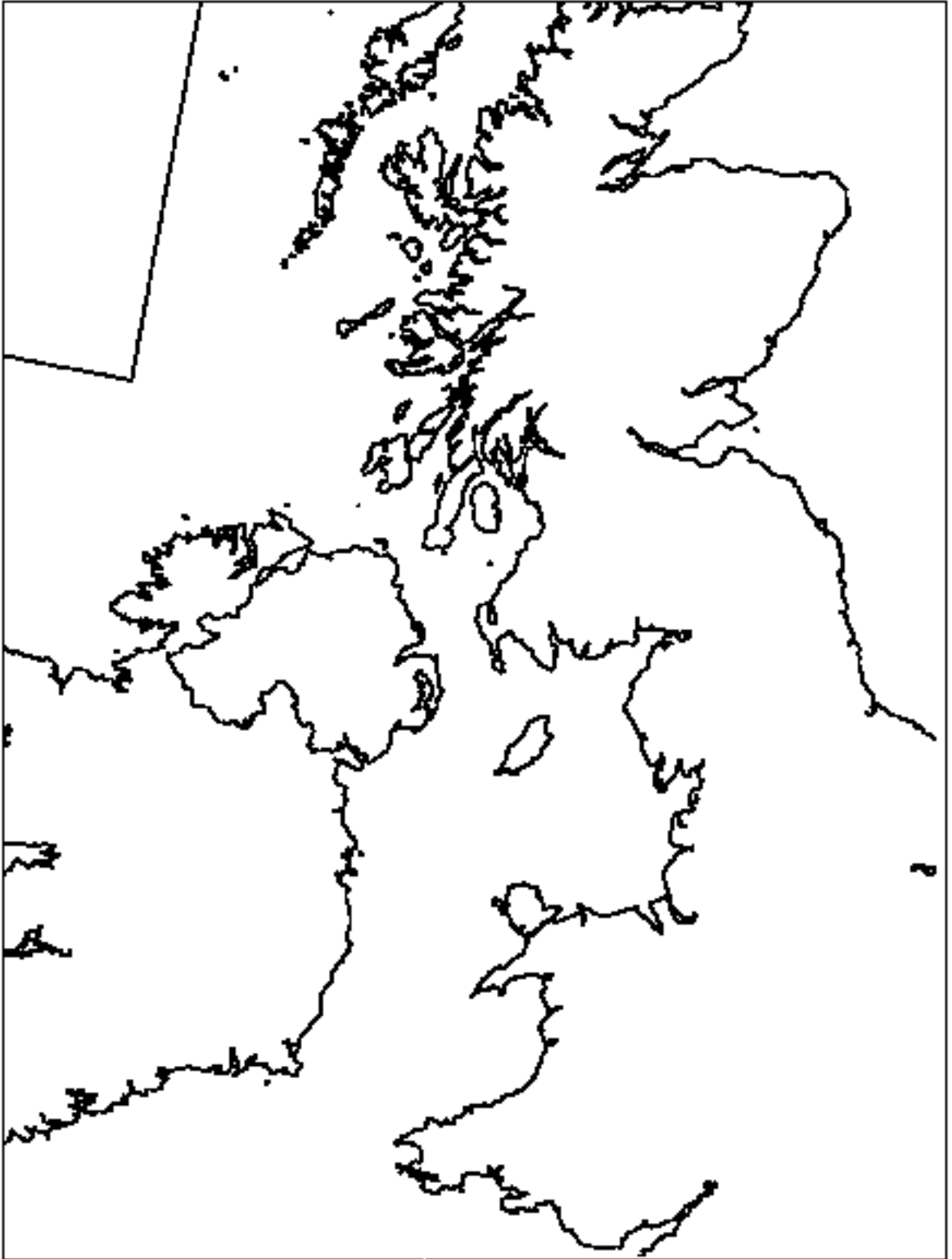


Figure 1: unmodified Political/Oceans layer (490K)

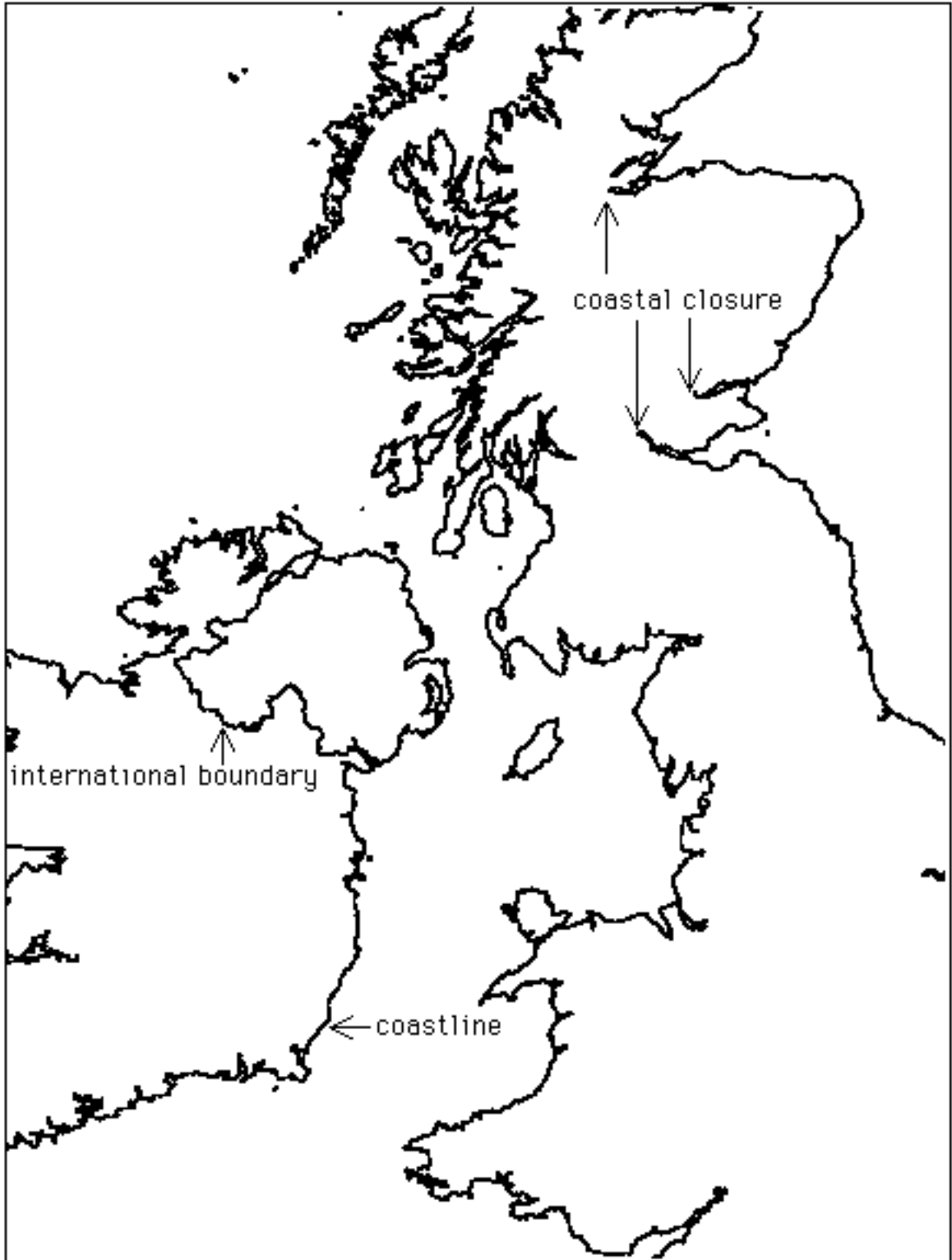


Figure 2: unmodified *land* (unfilled), simulating *coastline*, *de jure international boundaries* and *coastal closure shoreline* (400K)



Figure 3: coarsened *land* (unfilled), simulating coarsened *coastline*, *dejure international boundaries* and *coastal closure shoreline* (20K)



Figure 4: unmodified Drainage layer (779K)



Figure 5: unmodified *streams/rivers* (432K)

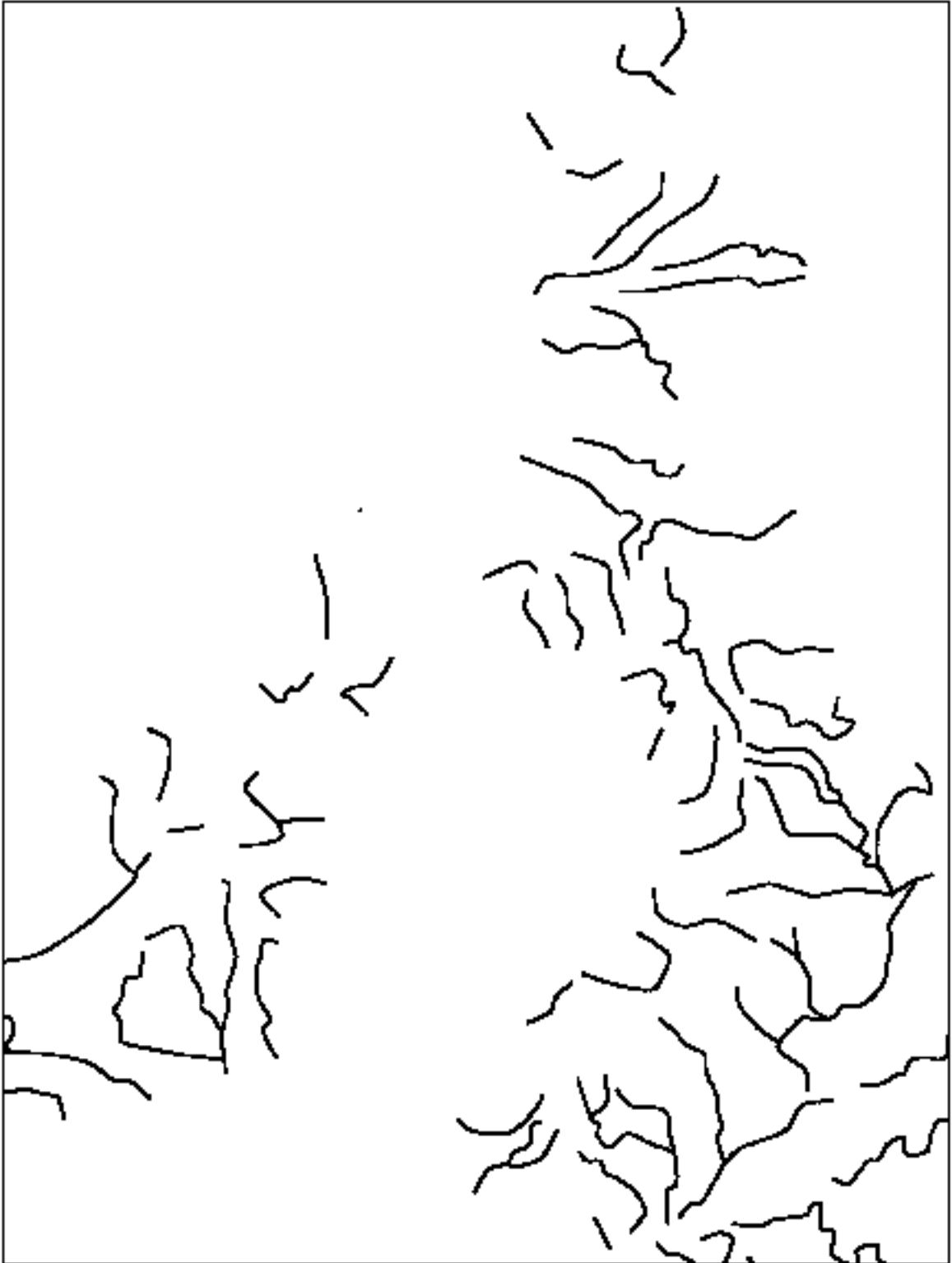


Figure 6: *streams/rivers* remaining after omission (17K)



Figure 7: unmodified *perennial inland water* (51K)



Figure 8: *perennial inland water* remaining after omission (22K)

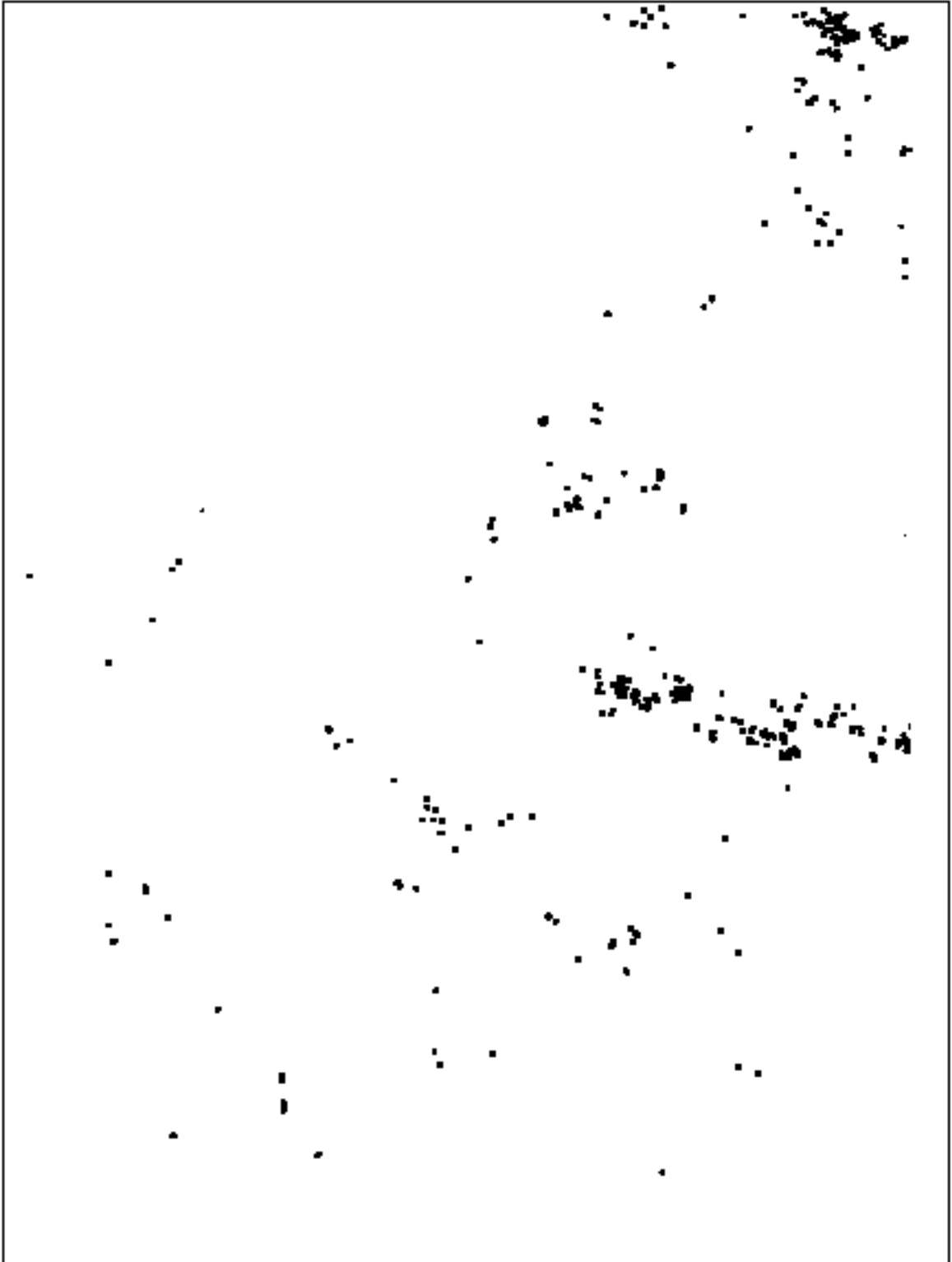


Figure 9: unmodified Drainage-Supplemental layer (60K)

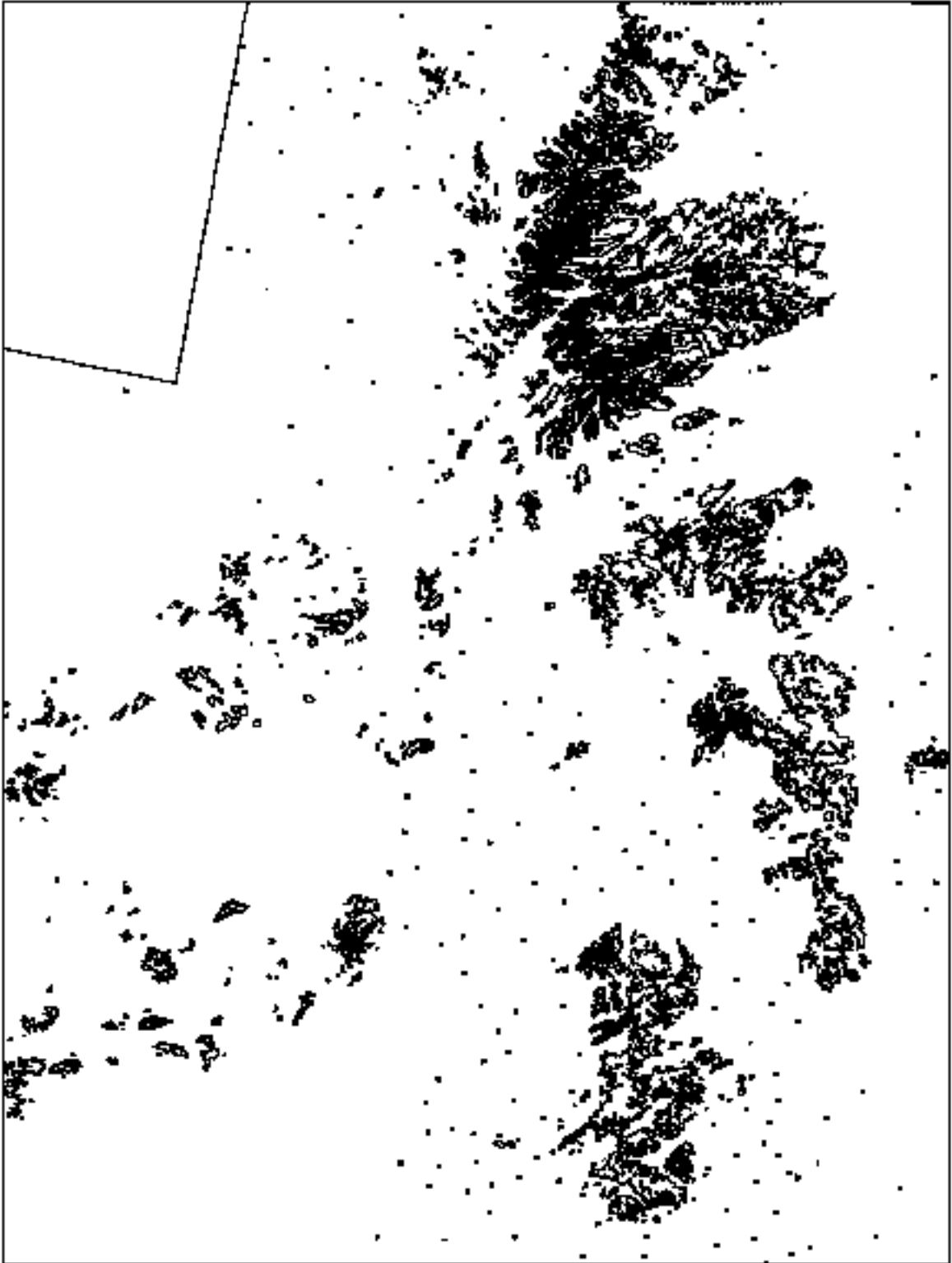


Figure 10: unmodified Hypsography layer (>1000K)



Figure 11: unmodified *1000 to 3000 feet* and *3000 to 7000 feet* (496K)



Figure 12: *1000 to 3000 feet and 3000 to 7000 feet* after WEEDDRAW operation (181K)



Figure 13: unmodified Hypsography-Supplemental layer (>1000K)



Figure 14: unmodified Vegetation layer (300K)



Figure 15: unmodified *vegetation* (285K)



Figure 16: *vegetation* after WEEDDRAW operation (120K)

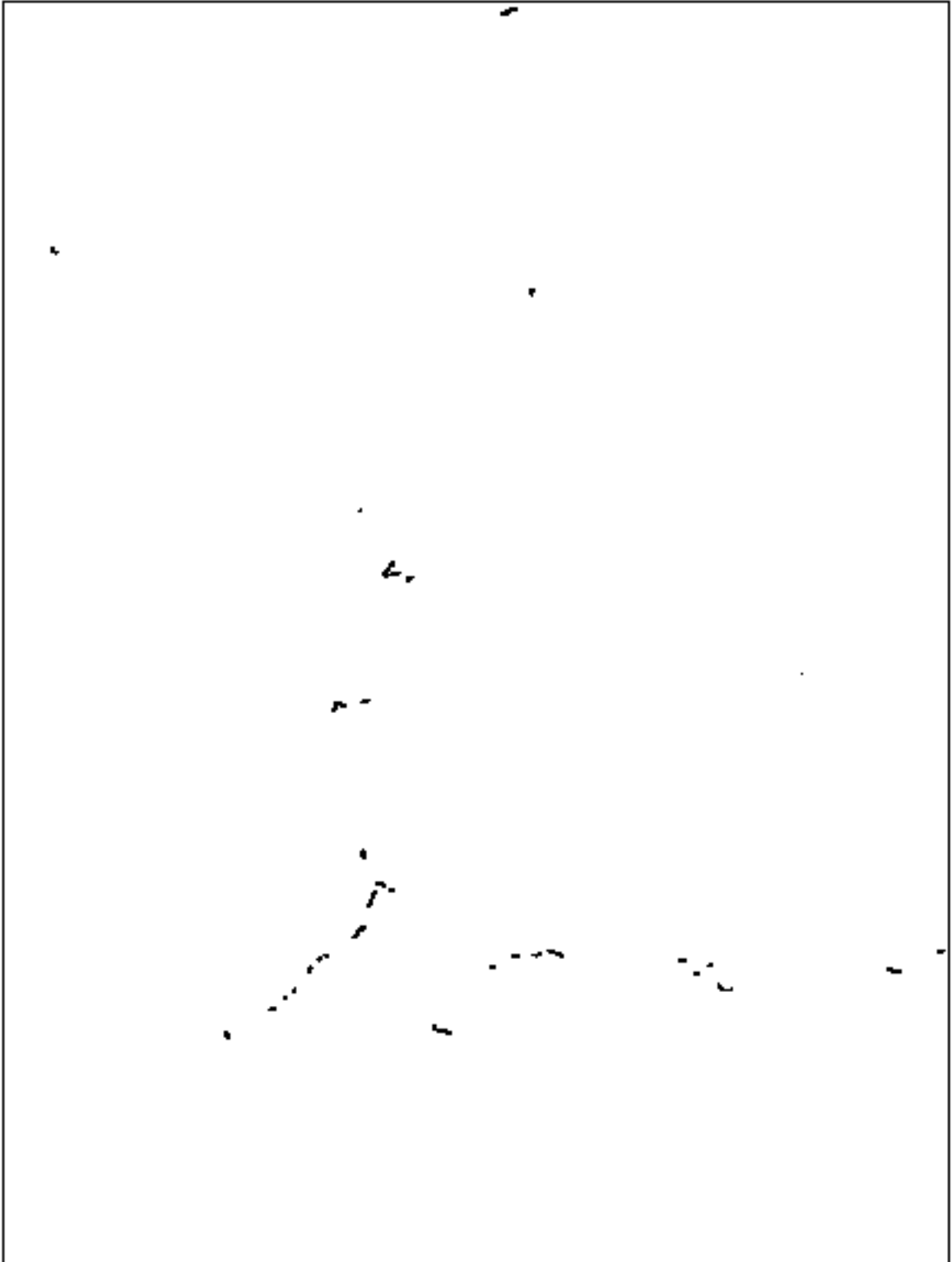


Figure 17: unmodified Physiography layer (3K)

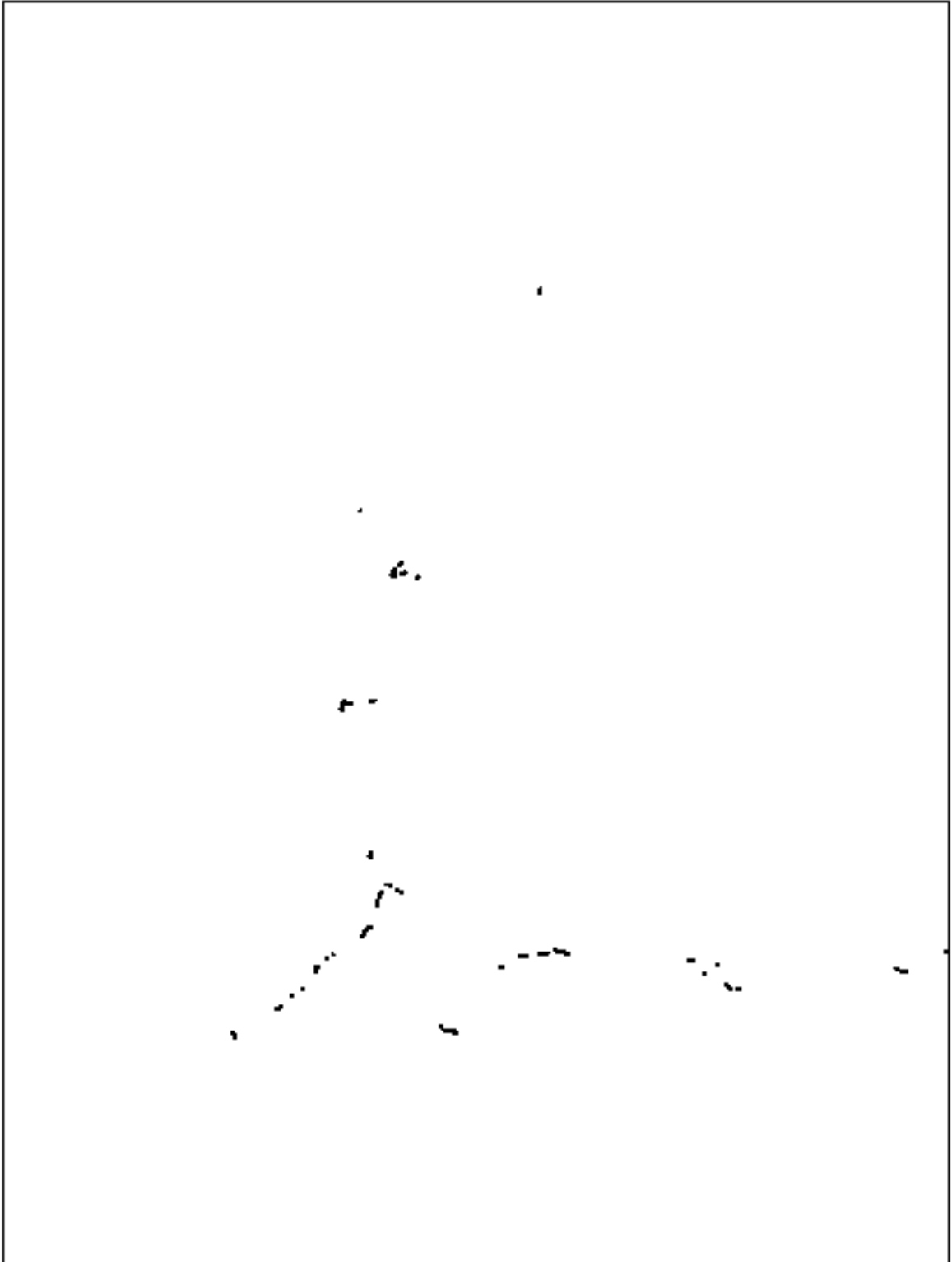


Figure 18: *escarpments, bluffs, cliffs, etc.* found along the coast (3K)

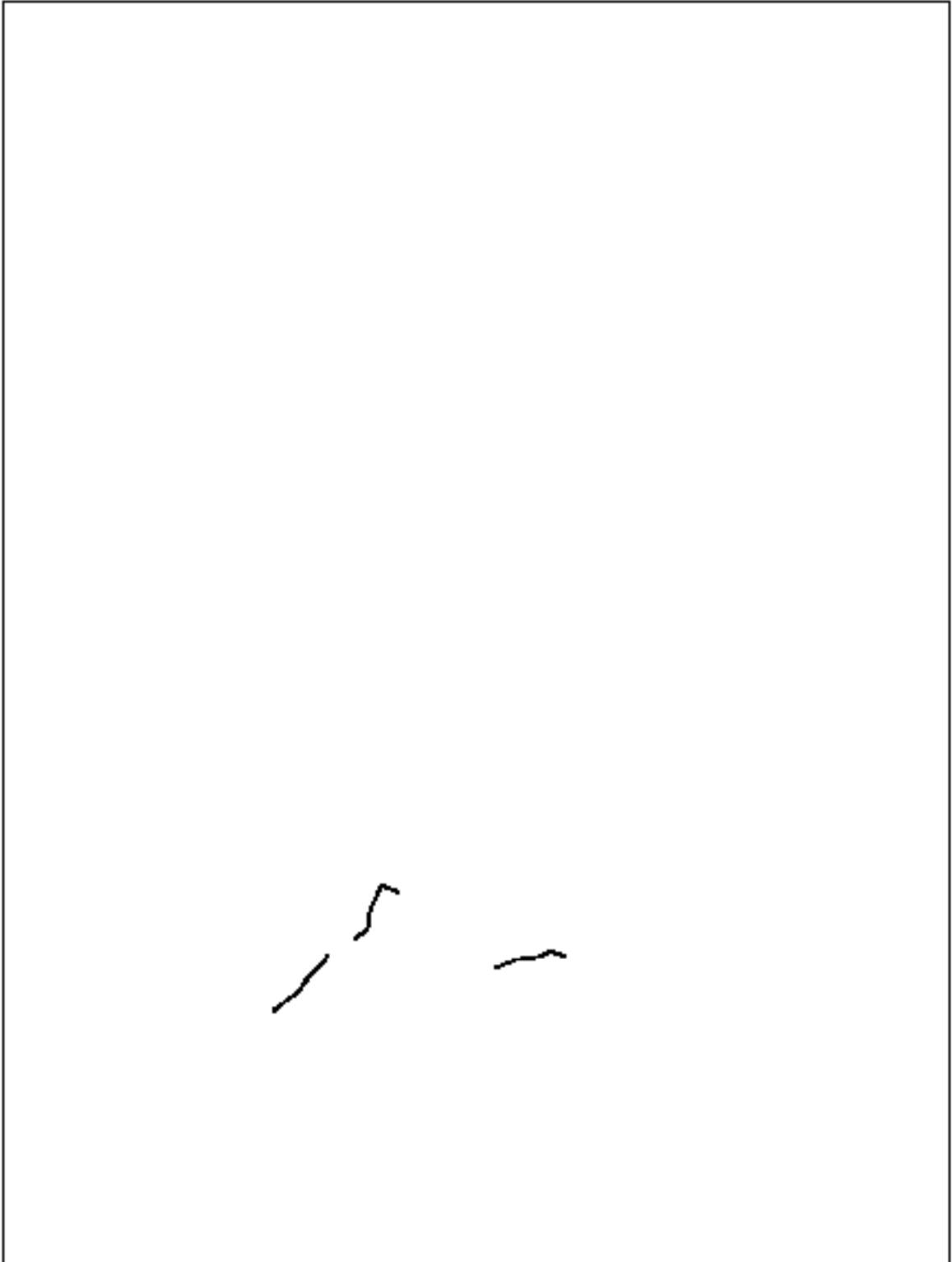


Figure 19: "escarpmented coastline" (<1K)



Figure 20: unmodified Ocean Features layer (115K)

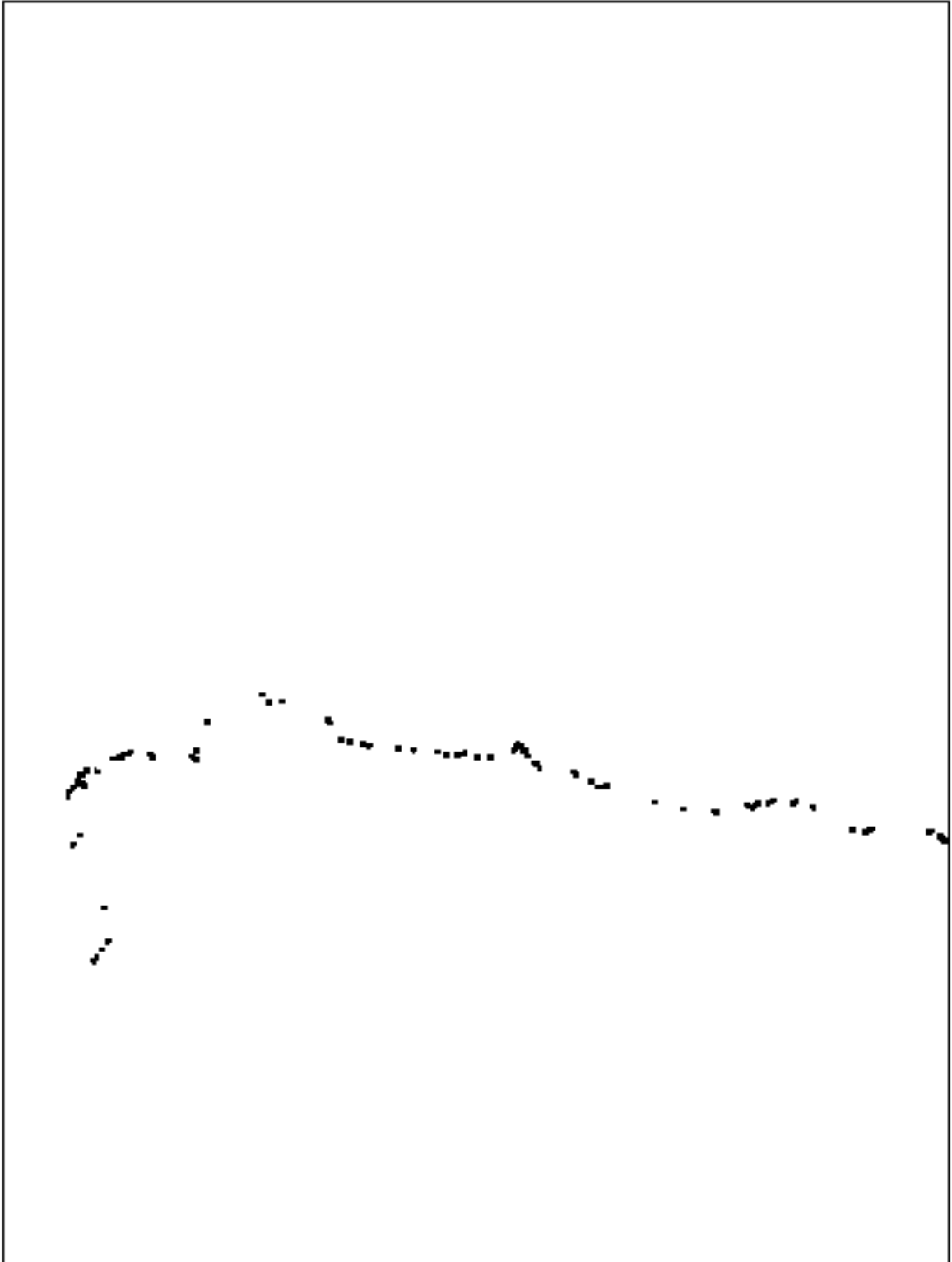


Figure 21: unmodified *rocks, isolated or awash* (11K)
(different geographic area coverage than figure 20)

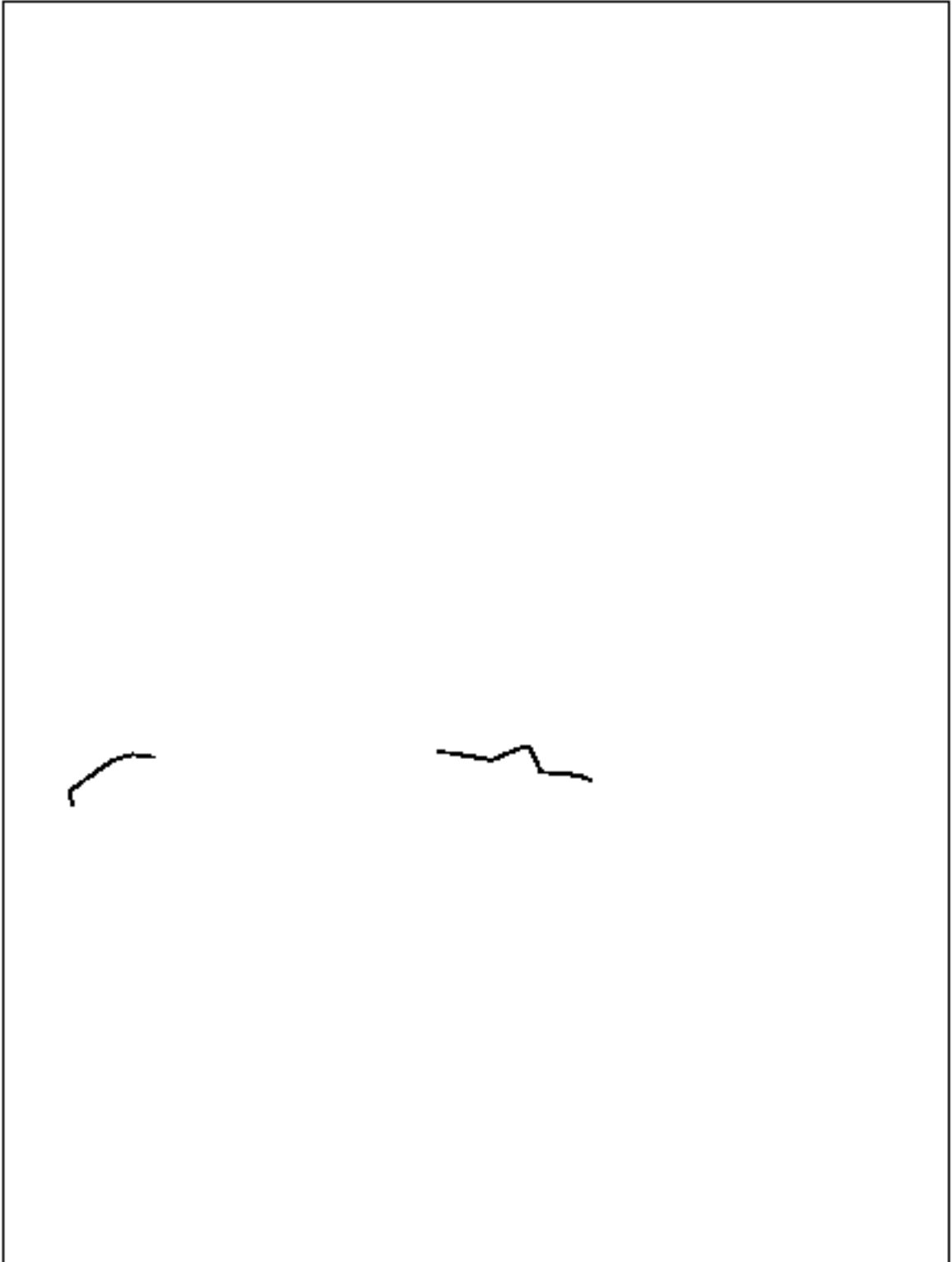


Figure 22: "rocky coastline" (<1K)
(different geographic area coverage than figure 20)



Figure 23: unmodified *reefs* (95K)



Figure 24: "reefed coastline" (2K)



Figure 25: unmodified Populated Places layer (337K)



Figure 26: unmodified *built-up areas* (176K)



Figure 27: *built-up areas* after WEEDDRAW operation (29K)



Figure 28: unmodified total point features in Populated Places layer (159K)

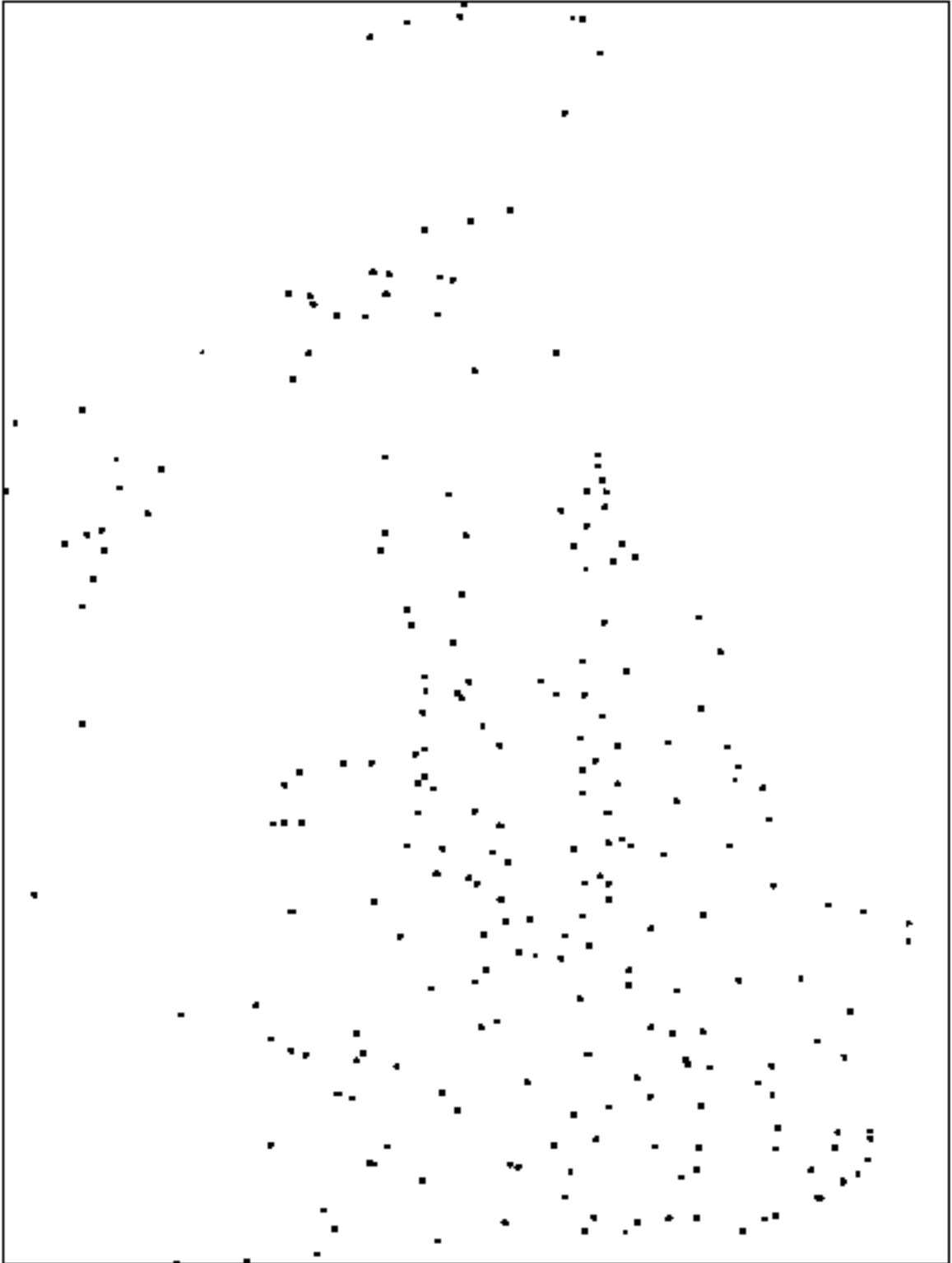


Figure 29: unmodified *populated places* (names in city tints) (29K)



Figure 30: unmodified Roads layer (460K)



Figure 31: unmodified *primary/secondary roads or highways* (381K)

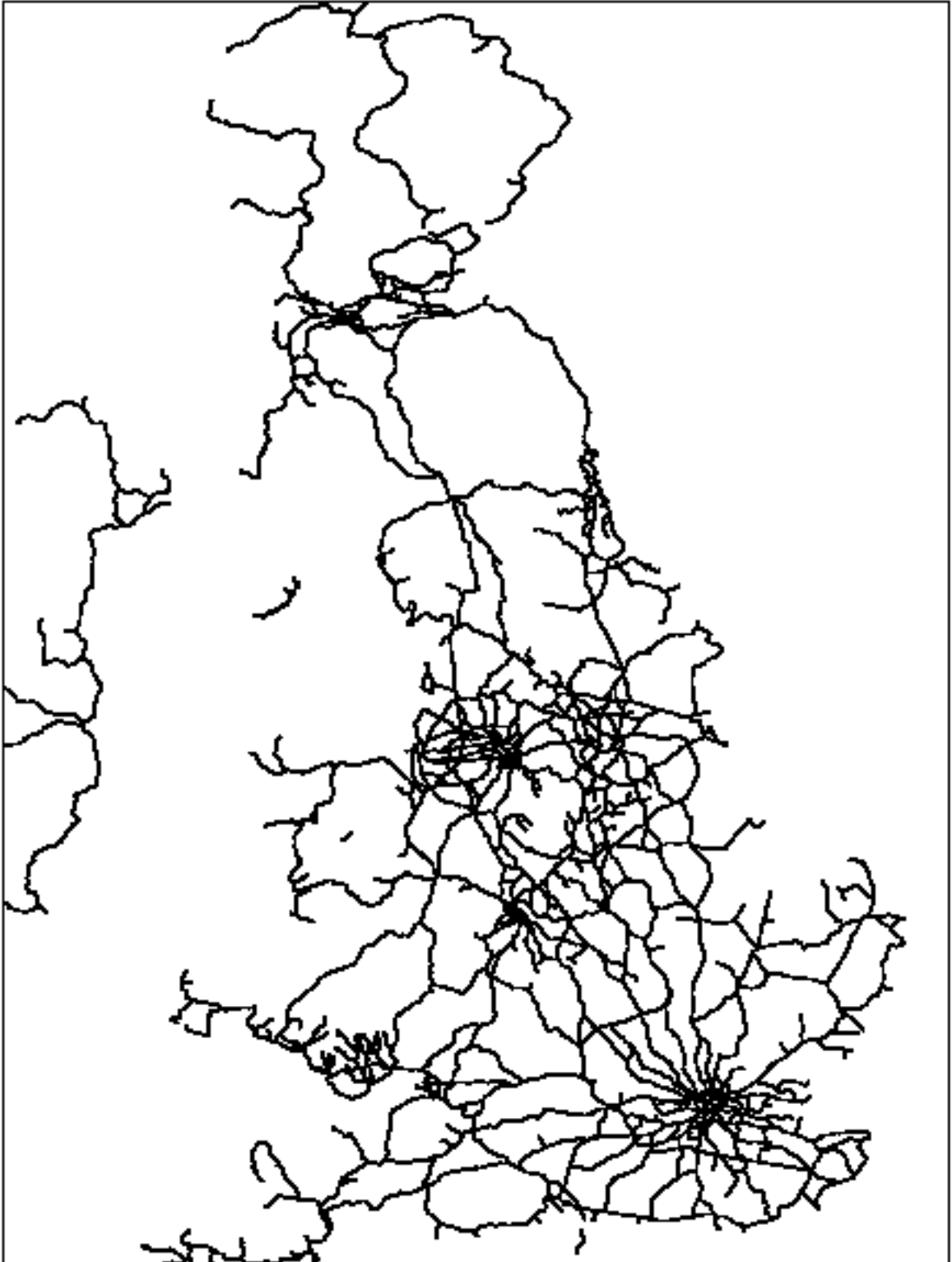


Figure 32: unmodified Railroads layer (148K)

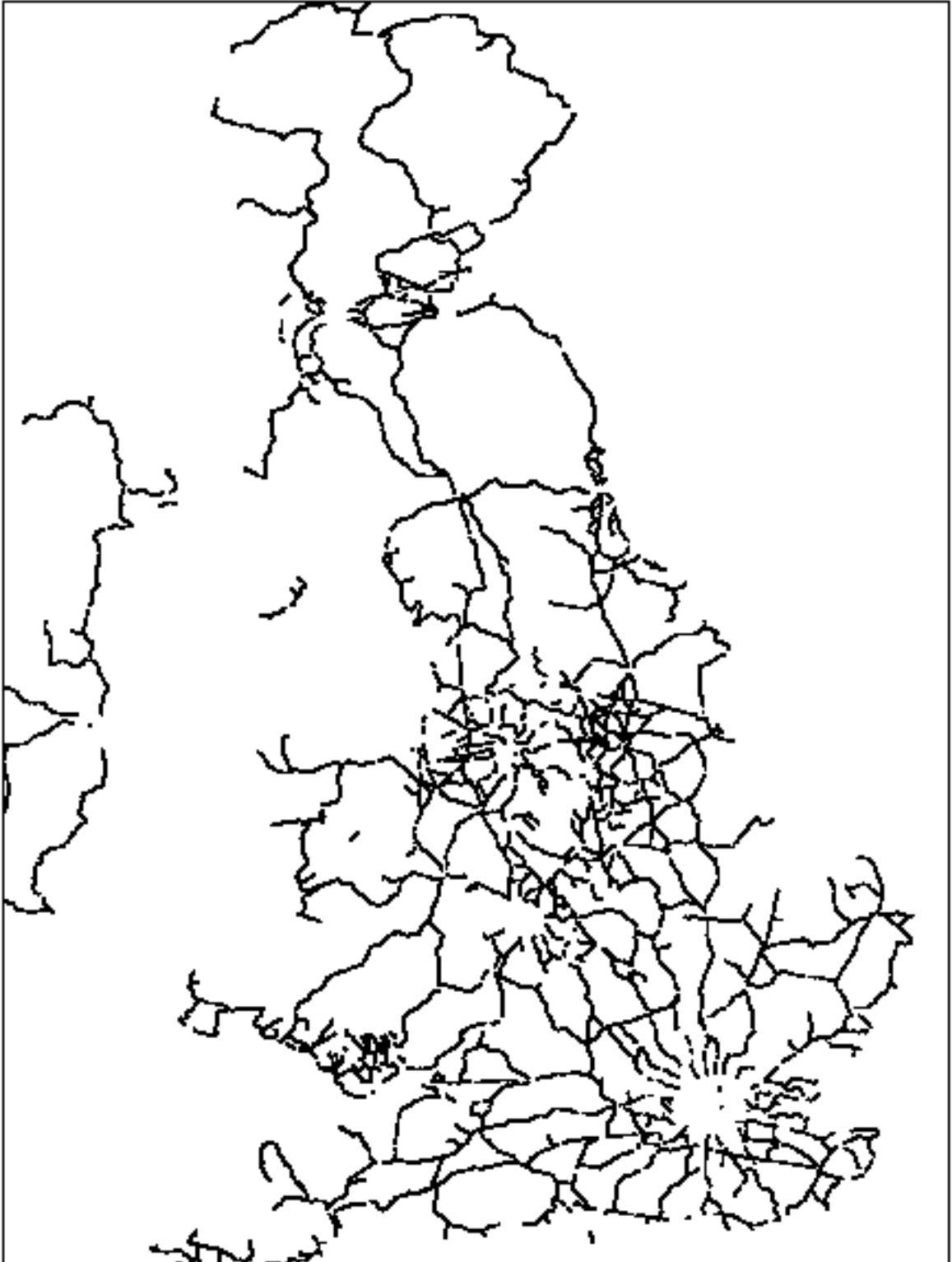


Figure 33: Railroads layer without *connectors* (116K)

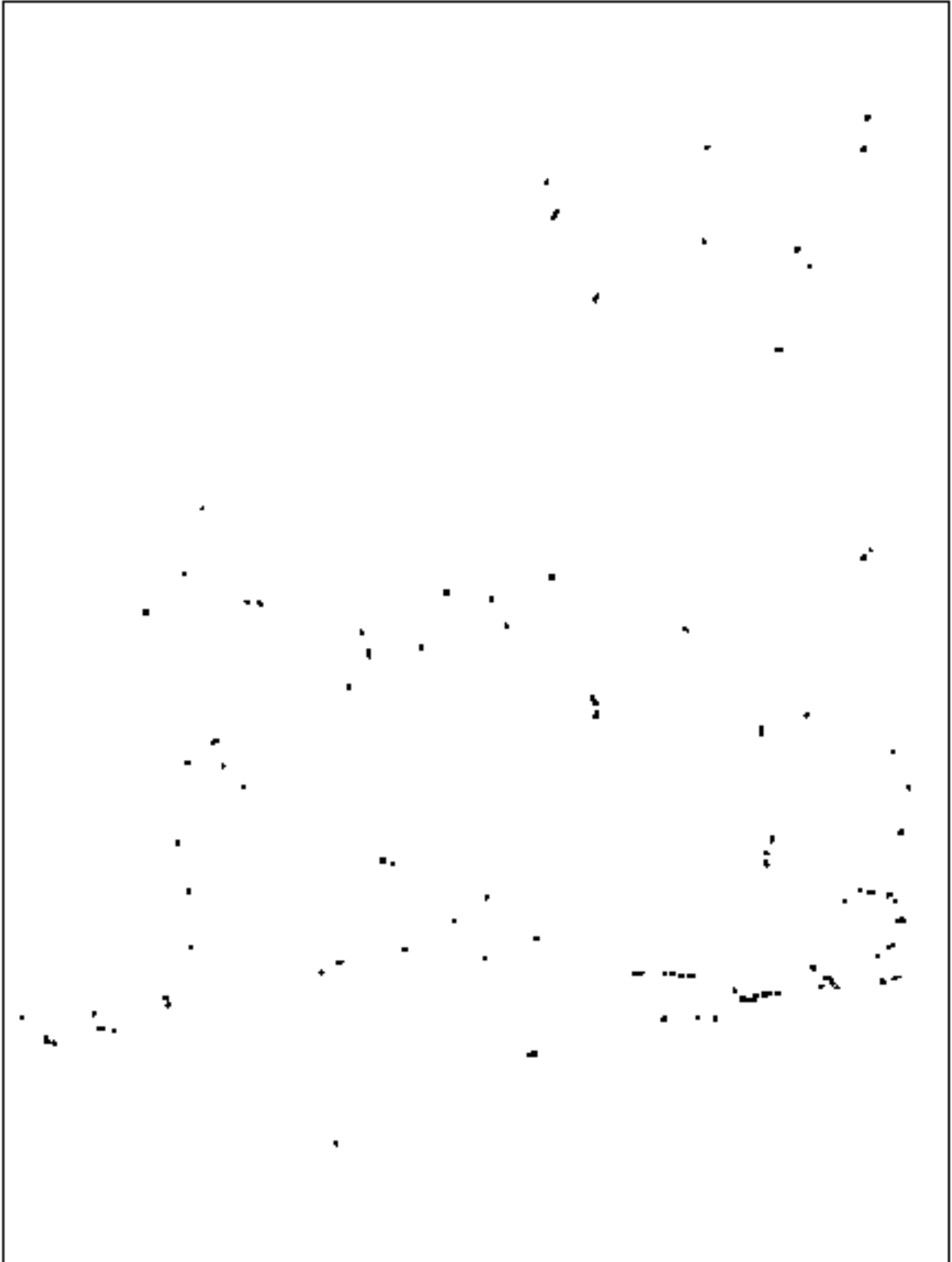


Figure 34: unmodified Transportation Structure layer (8K)

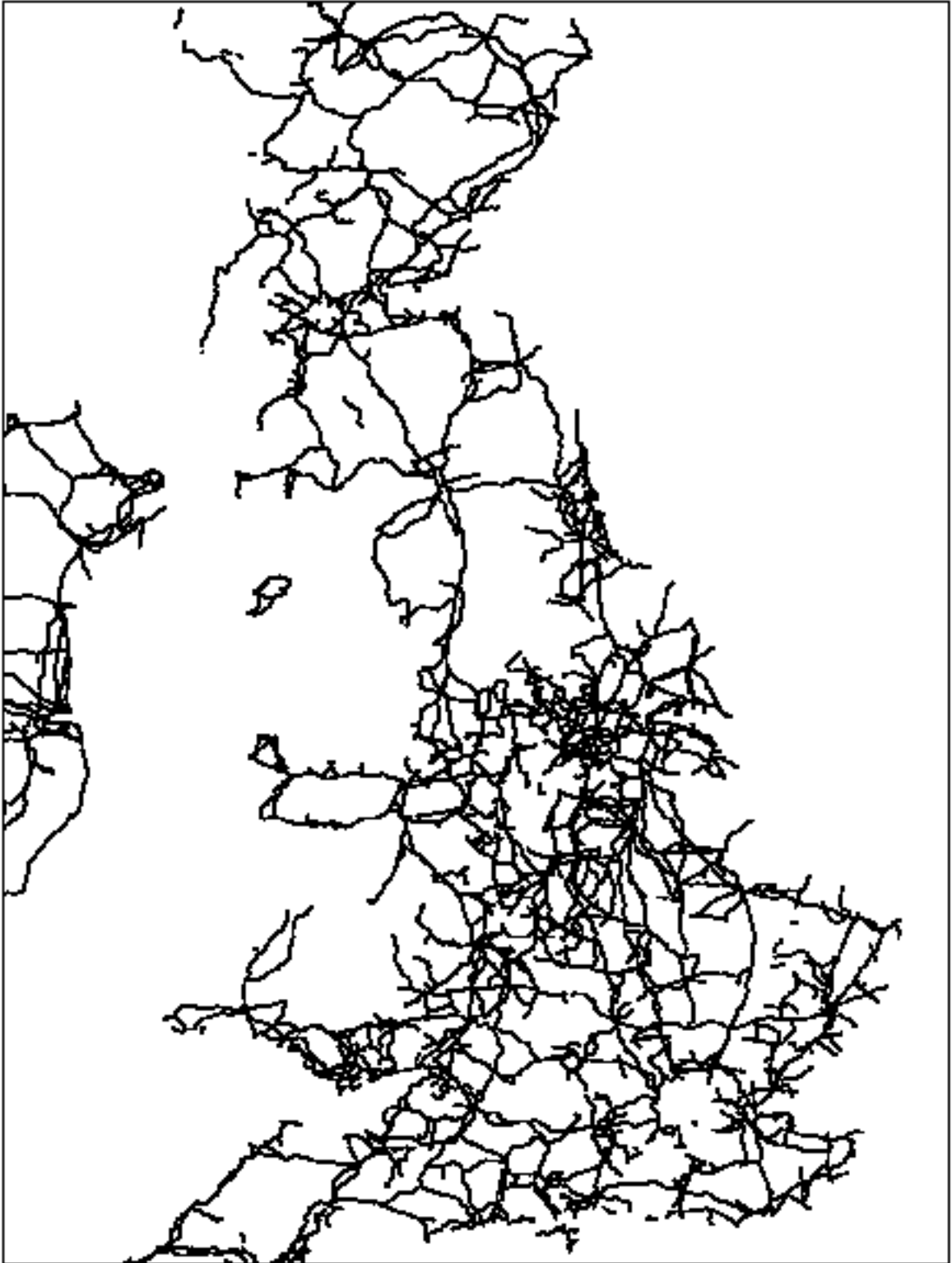


Figure 35: unmodified Utilities layer (191K)

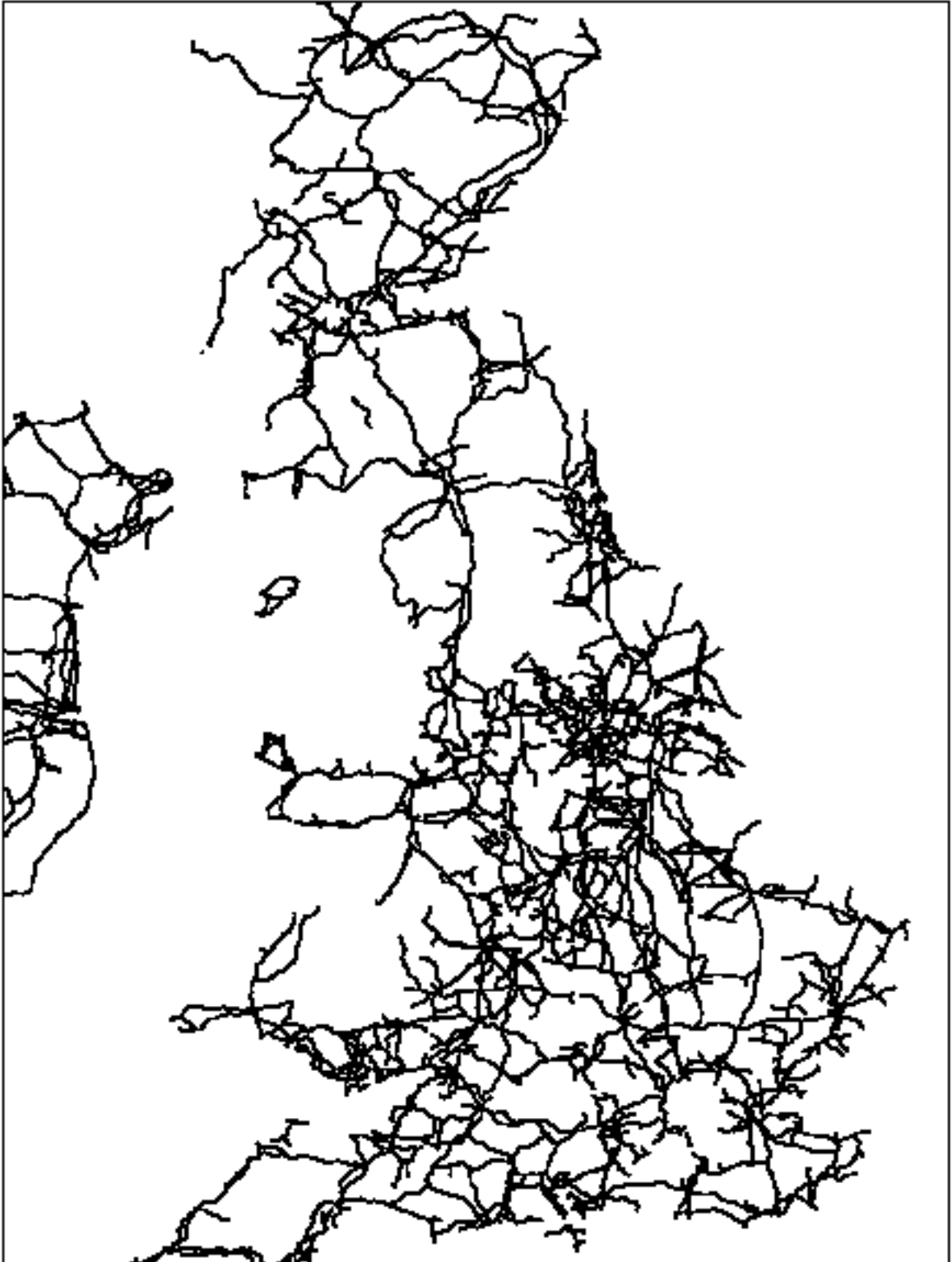


Figure 36: *power transmission lines* remaining after omission (190K)



Figure 37: unmodified Land Cover layer (183K)



Figure 38: unmodified *undifferentiated wetlands* (125K)



Figure 39: *undifferentiated wetlands* after WEEDDRAW operation (76K)



Figure 40: unmodified *unconsolidated materials* (49K)



Figure 41: *unconsolidated materials* after WEEDDRAW operation (21K)

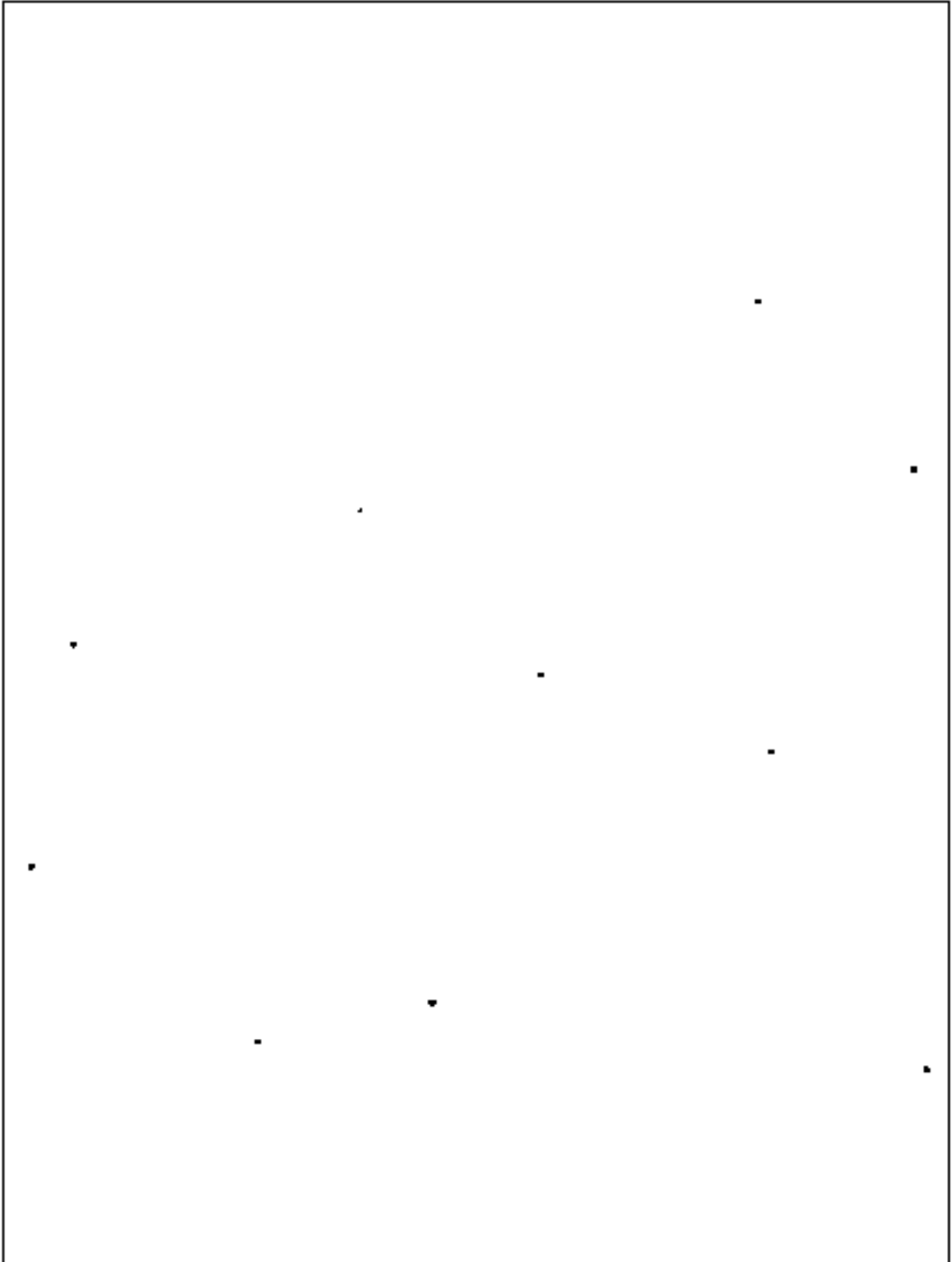


Figure 42: unmodified Landmarks layer (3K)

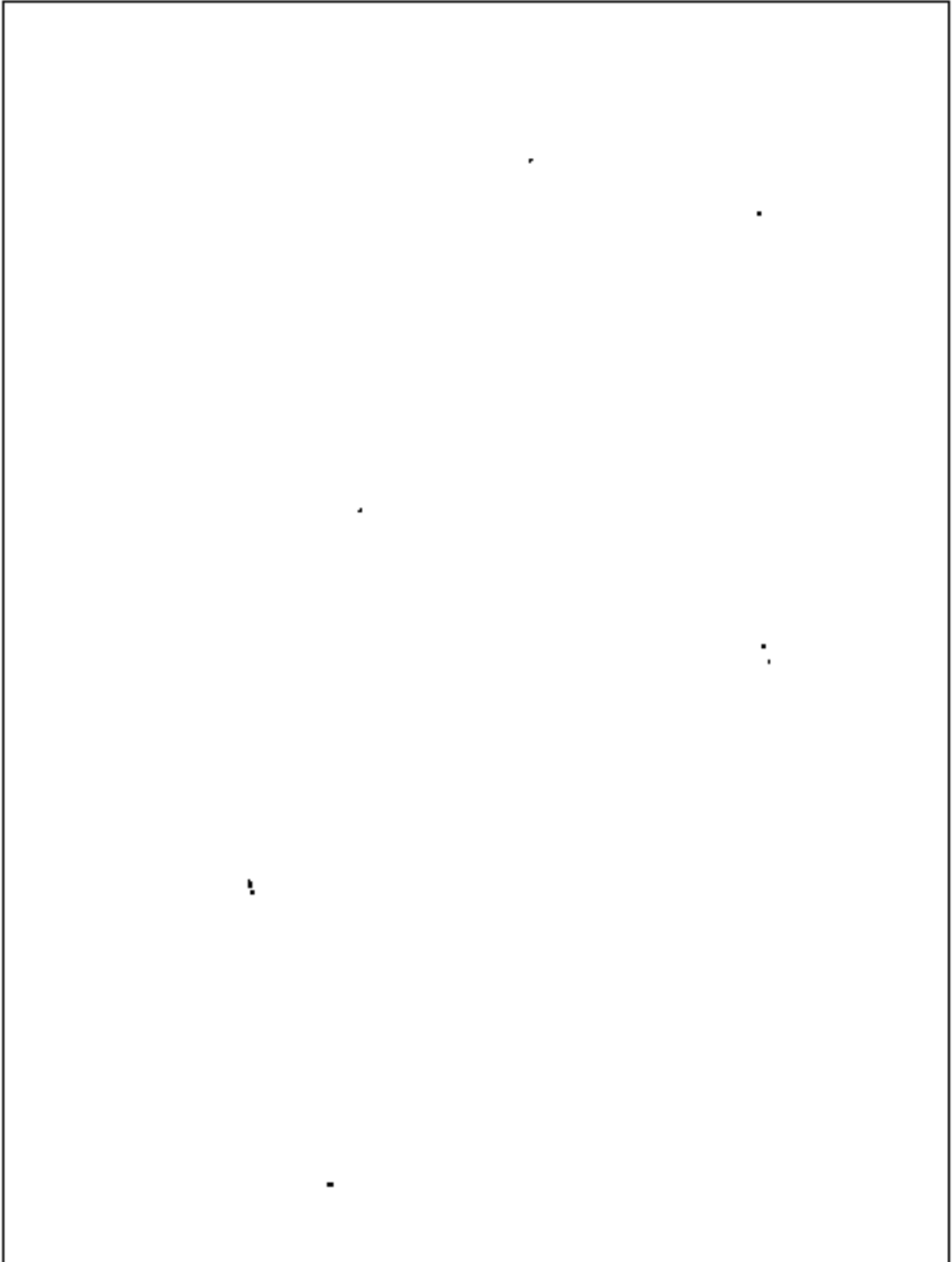


Figure 43: unmodified General Culture layer (1K)



Figure 44: unmodified Aeronautical layer (19K)

Appendix B: DCW Features in the Study Area

Following is a list of the features present in the special ARC/INFO version of DCW Prototype 4 used for this project, arranged by layer. The letter codes define the status of the feature(s) or layer with respect to the new "gap-fill" database:

- (D) not included in new database
- (G) generalized before placed in new database
- (R) retained without modification in new database

Appendix C: Generalization Operator Definitions*

SELECT: This is a special operator which must precede all others. It is required to initialize the composition of a graphic view of the database which can then be displayed on a monitor or as hardcopy output. The user is informed of information stored in the database and from this they may select items by theme, feature type, or instance. This operation allows the user to explicitly choose only desired items. For example, the user may select the theme roads, in which case all roads in the selected geographic area will be extracted for display. The user may also be more specific and select only Interstate Highways or to be most specific, select only Interstate 95 for example.

OMIT: Once items have been selected for display, the omit operator allows removal of objects. These objects are only removed from the display list and not from the database. As with the SELECT operator, individual objects may be removed or objects may be removed by theme, feature type, or conflict type. For example, the OMIT operator could be used to remove all objects which were too small.

COARSEN: This operator removes line spatial detail (crenellations from a line). This operator could be applied to objects stored in the database with a high level of spatial detail, and which the user wishes to display in less detail. The operator works primarily on metric detail, but may change topology of objects (for example, by removing a small island from a lake whose boundary has been coarsened). The user need not specify parameters for this operator. They only need select the object or objects to be coarsened and apply the operator. The operator uses the minimum thresholds defining areas too small, items too close or items too narrow which have been computed for the selected

*adapted from Beard and Mackaness (1991)

scale or format. The resulting representation is therefore appropriate to the selected scale. This operator can be applied to individual objects, themes or feature types.

COLLAPSE: The collapse operator substitutes a 1D or 0D representation for a 2D representation. This operator could be applied to objects stored in the database as areas, but which a user wishes to display as points or lines. This operator must be preceded or succeeded by a symbol change. COLLAPSE resolves the legibility problem of items being too close, or COLLAPSE followed by a change in symbol width could resolve the problem of items being too small or too narrow.

COMBINE: The combine operator simplifies a spatial representation by merging objects which are nearby in space into a single new object. For example a cluster of small islands may be combined to form a larger island. The operator applies only to two or more selected objects and the result is always one new object. Thus COMBINE is strictly a localized operator. This operation must be preceded or succeeded by the CLASSIFY operator so that the resulting object is properly identified. COMBINE resolves items being too small or too close.

AGGREGATE: This operator is similar to COMBINE but merges objects which are adjacent rather than those with intervening spaces. CLASSIFY must precede this operator as well. The aggregate operator can be applied globally by theme or by feature class.

CLASSIFY: This operator allows individual objects, feature types or themes to be assigned to a new class. The classification may be based on shared attribute characteristics of objects. The user or systems selects a set of objects and assigns a new

class label (e.g. For all objects with attribute D, Class = M). A symbol change must follow this operation, and when the new symbol is assigned, all objects assigned to the new class inherit the symbol. This operator does not directly resolve conflicts but is required as a supporting operation for operators which change the nature of an object (i.e. COMBINE and AGGREGATE).

EXAGGERATE: The exaggerate operator expands the size or width of objects. It can be applied by theme, feature type, instance or conflict type. The operator expands the object to meet the minimum threshold for legibility and therefore requires no parameter specification by the user. For a line or point representation, the width or radius is expanded. This can be accomplished by redimensioning a symbol. For an area, the operation performs a localized scale increase.

DISPLACE: This operator is applied locally to two or more objects which are too close or overlapping.

References

- Beard, M. K. and W. A. Mackaness, 1991. Generalization operations and supporting structures. **Proceedings, Tenth International Symposium on Computer-Assisted Cartography (Auto-Carto 10)**, pp. 29-45.
- Brassel, K. E. and R. Weibel, 1988. A review and conceptual framework of automated map generalization. **International Journal of Geographic Information Systems**, vol. 2, pp. 229-244.
- Buttenfield, B. P., 1989. Scale-dependence and self-similarity in cartographic lines. **Cartographica**, vol. 26(1), pp. 79-100.
- Buttenfield, B. P., 1991. "A Rule for Describing Line Feature Geometry". Chapter 9 in **Map Generalization: Making Decisions for Knowledge Representation**. (Buttenfield, B. P. and R. B. McMaster Editors) London: Longman, pp. 150-171.
- Buttenfield, B. P. and J. S. DeLotto, 1989. **Multiple Representations**, NCGIA Technical Paper 89-3. Santa Barbara, CA: NCGIA.
- Buttenfield, B. P. and D. M. Mark, 1991. "Expert Systems in Cartographic Design". Chapter VII in **Geographic Information Systems: The Computer and Contemporary Cartography**, (D. R. F. Taylor, Editor). Oxford: Pergammon Press, Ltd., pp. 129-150.

- Douglas, D. and T. K. Peucker, 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. **The Canadian Cartographer**, vol. 10, pp.112-122.
- DMA (Defense Mapping Agency), 1990. **Digitizing the Future**, third edition. DMA stock number DDIPDIGITALPAC.
- ESRI (Environmental Systems Research Institute, Inc.), 1989. **ARCPLLOT User's Guide**. Redlands, CA: Environmental Systems Research Institute, Inc.
- ESRI (Environmental Systems Research Institute, Inc.), 1990. **Draft Product Specification for the Digital Chart of the World**. Contract Data Requirements List document number B002, fulfilling DMA contract number DMA600-89-C-0023.
- Mackness, W. A. and P. F. Fisher, 1987. Automatic recognition and resolution of spatial conflicts in cartographic symbolization. **Proceedings, Eighth International Symposium on Computer-Assisted Cartography (Auto-Carto 8)**, pp. 709-718.
- Mark, D. M., 1989. Conceptual basis for geographic line generalization. **Proceedings, Ninth International Symposium on Computer-Assisted Cartography (Auto-Carto 9)**, pp. 68-77.
- McMaster, R. B., 1987. Automated line generalization. **Cartographica**, vol. 24, pp. 74-111.

- McMaster, R. B. and K. S. Shea, 1988. Cartographic generalization in a digital environment: a framework for implementation in a geographic information system. **Proceedings, GIS/LIS '88**, vol. 1, pp. 240-249.
- Muller, J. C., 1990a. The removal of spatial conflicts in line generalization. **Cartography and Geographic Information Systems**, vol. 17, pp. 141-149.
- Muller, J. C., 1990b. **Generalization of Spatial Data Bases**. Draft manuscript copyrighted by International Institute for Aerospace Surveys and Earth Sciences (ITC), Enschede, The Netherlands.
- Rand McNally, 1968. **World Atlas, Imperial Edition**.
- Shea, K. S. and R. B. McMaster, 1989. Cartographic generalization in a digital environment: when and how to generalize. **Proceedings, Ninth International Symposium on Computer-Assisted Cartography (Auto-Carto 9)**, pp. 56-67.
- Weibel, R., 1987. An adaptive methodology for automated relief generalization. **Proceedings, Eighth International Symposium on Computer-Assisted Cartography (Auto-Carto 8)**, pp. 42-49.
- Zoraster, S., D. Davis and M. Hugus, 1984. **Manual and Automated Line Generalization and Feature Displacement**. U. S. Army Engineer Topographic Laboratories report number ETL-0359.

AERONAUTICAL LAYER: (R)

point feature:
airport

DRAINAGE LAYER:

point feature: (R)
dams

line features:
streams/rivers (G)
inland shorelines (D)
canals, aqueducts, etc (R)
connector (D)
none (D)

area features:
perennial inland water (G)
non-perennial inland water (R)
none (inland water island) (D)

DRAINAGE-SUPPLEMENTAL LAYER:
(D)

point features:
small lakes
small inland water islands

GENERAL CULTURE LAYER:

line feature:
prominent fences (D)

HYPSOGRAPHY LAYER:

point features: (R)
spot elevation
spot elevation, location doubtful

line features: (D)
closed land contour
connector
none

area features: (G)
0 to 1000 feet
1000 to 3000 feet
3000 to 7000 feet
7000 to 11,000 feet
11,000 feet and above

HYPSOGRAPHY-SUPPLEMENTAL
LAYER: (D)

line feature:
partial intermediate or auxiliary contour

LAND COVER LAYER:

point feature: (R)
mines

area features:

salt pans (R)
unconsolidated materials (G)
undifferentiated wetlands (G)
none (D)

LANDMARKS LAYER: (R)

point features:
building
dam
factory
lighthouse
monument
tanks
tower

OCEAN FEATURES LAYER:

point features:
rocks, isolated or awash (G)
exposed wrecks (R)

line feature: (G)
reefs

PHYSIOGRAPHY LAYER:

line features:
levees, dikes, eskers (R)
escarpments, bluffs, cliffs, etc (G)

POLITICAL/OCEANS LAYER:

point feature: (D)
small island

line features:
dejure international boundaries (G)
coastal closure shoreline (G)
coastline (G)
none or unknown (module border) (D)

area features: (R)
land
ocean

POPULATED PLACES LAYER:

point features:

populated places (D)

populated places (names in city tints) (R)

area features:

built-up areas (G)

none (D)

RAILROADS LAYER: (R)

line features:

single track railroads

multiple track railroads

connectors

ROADS LAYER: (R)

line features:

dual lane (divided) highways

primary/secondary roads or highways

connectors

TRANSPORTATION STRUCTURE
LAYER: (D)

point features:

road structures

railroad structures

line features:

road structures

railroad structures

UTILITIES LAYER: (G)

line feature:

power transmission lines

VEGETATION LAYER:

area features:

vegetation (G)

hole in vegetation/none (D)