Jingfen Sheng,[1] John P. Wilson, and Ning Chen
Department of Geography, University of Southern California Los Angeles, California 90089-0255

Joseph S. Devinny and Jaime M. Sayre
Department of Civil and Environmental Engineering, University of Southern California Los Angeles, California 90089-2531

Abstract: The purpose of this paper is to evaluate the "fitness for use" of the high-resolution National Hydrography Dataset (NHD) for regional watershed assessments. The Nature Conservancy's GIS tools were combined with various data sources and techniques to identify and fix four types of problems in the high resolution NHD—dangling streams, attribute errors, flow divergences, and duplicate streams. This effort generated a single-line natural flow network with correct arc-node topology, no dangling streams (with the exception of streams terminating in low relief basins), and few, if any, attribute designation problems. The revised NHD network improved on the original representation of the stream network and accompanying sub-watersheds in several ways.

INTRODUCTION

Geographic information system (GIS) tools and datasets are needed to assess historical trends and current conditions in regional watershed conservation and stream restoration projects (Groves et al., 2000; Hargrove et al., 2002; Roni et al., 2002). The National Hydrography Dataset (NHD) is perhaps the most complex of these datasets. It supports regional hydrologic analysis and has been integrated with a variety of GIS-based analytical tools such as the Nature Conservancy's freshwater ecological classification tools (TNC GIS tools) for performing regional-scale ecological assessments of watersheds (e.g. Wittner and McKee, 2000).

The NHD incorporates numerous choices in terms of the spatial representation of water entities on the Earth's surface which, in turn, influence the quality or "fitness for use" of these data. This "fitness for use" may vary with the choice of analytical tools and landscape setting (i.e., study area). We therefore set out to evaluate the

[1]sheng@usc.edu

GIScience & Remote Sensing, 2007, **44**, No. 3, p. 283–304.
Copyright © 2007 by V. H. Winston & Son, Inc. All rights reserved.

"fitness for use" of the NHD for regional watershed assessments. The TNC GIS tools were combined with various data sources and techniques to characterize the problems and modifications that should be made to the NHD in order to perform watershed assessments in large metropolitan areas like those found in southern California. The revised NHD stream network was used to delineate watershed boundaries and generate hydrologic attributes to examine the impacts of NHD data representation on the results of the TNC watershed assessments. The results of using the original and revised NHD show that the spatial representation of a watershed really matters in generating watershed-based attributes.

The remainder of this paper is divided into three sections. The first describes the key features of the NHD and TNC GIS tools and how they can be combined to perform regional watershed assessments. The second section describes our experiences using the TNC GIS tools and NHD in southern California. It briefly describes the study area and goals and then discusses the problems (errors) uncovered utilizing these tools and datasets. The GIS procedures and reference datasets that were deployed to solve various problems prior to watershed assessment are described and then the San Gabriel River watershed is utilized to illustrate the consequences of using the original and revised NHD for hydrologic parameter generation. The final section discusses the broader implications of our approach and future work that will advance these applications.

THE BASIC BUILDING BLOCKS

The NHD is a comprehensive set of digital spatial data that encodes information about naturally occurring and constructed bodies of water, the paths through which water flows, and related entities (EPA and USGS, 2000). The initial 1:100,000-scale, medium-resolution NHD was developed by integrating the U.S. Geological Survey (USGS) 1:100,000-scale Digital Line Graph (DLG) hydrography data with reach-related information from the U.S. Environmental Protection Agency (EPA) Reach File Version 3.0 (RF3). This product was completed and made available for the contiguous United States and Hawaii in 2001. A high-resolution NHD at scales of 1:24,000 to 1:12,000 that retains the linkages to the medium-resolution NHD reach delineations and associated information (Dewald and Roth, 1998; Gadd, 2002) is currently under development (see http://nhdgeo.usgs.gov/viewer.htm for latest status map).

The coverages and relational tables for each 8-Hydrologic Unit Code (HUC) watershed are organized in a single workspace. Each NHD workspace is implemented in Arc/Info as a georelational model that represents geographic features in spatial and descriptive terms in three coverages, labeled NHD, NHDPT, and NHDDUU, respectively. The NHD coverage stores the stream network and water body features, and incorporates route, region, and node subclasses. The NHDPT coverage contains all the point features such as gauging stations, wells, and dams. The locations, shapes, and/or extents of up to 52 different hydrologic features may be delineated and represented in these NHD and NHDPT coverages (see Fig. 1 for the most frequent and important features included in these coverages). A third coverage, NHDDUU, contains metadata information that documents the sources and methods used in the production of these datasets. The relational tables list attributes that describe feature

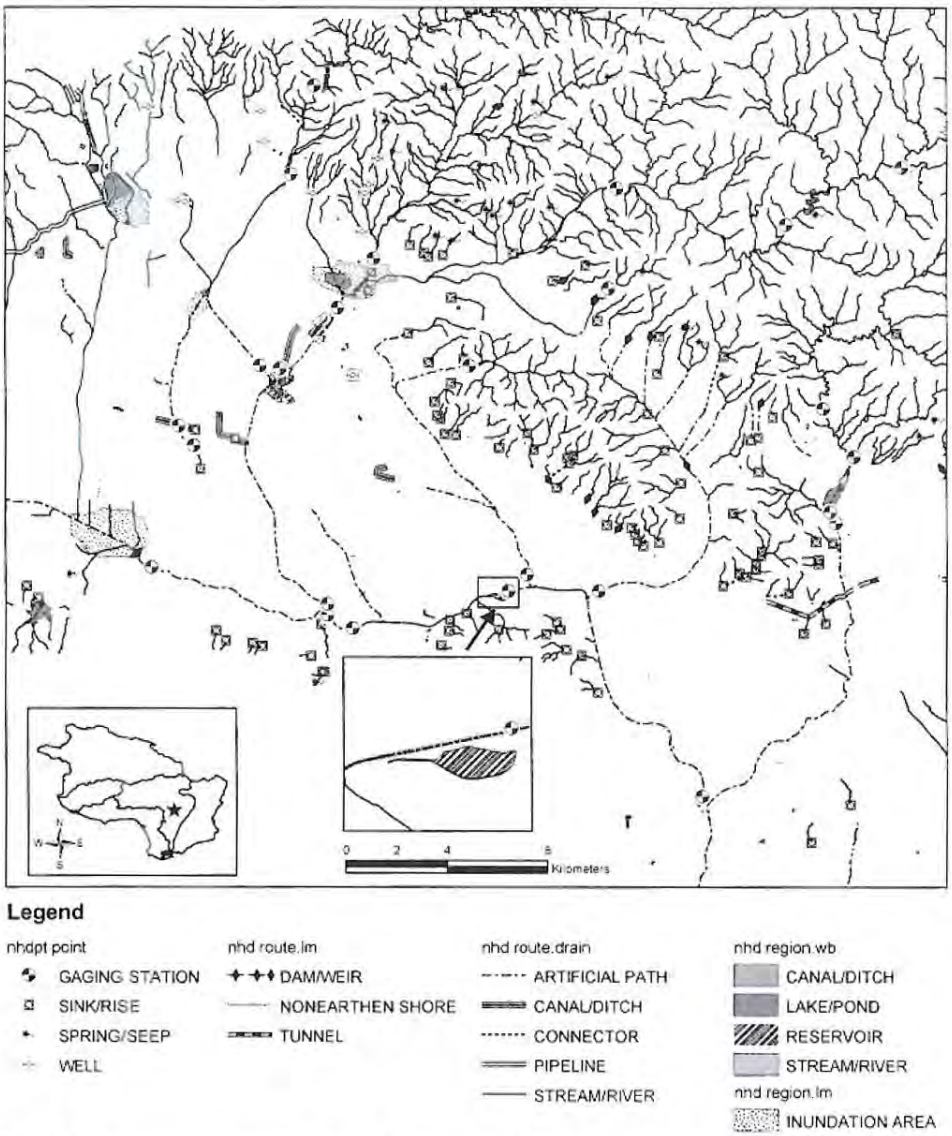

Fig. 1. Features represented by points, lines, or polygons in high-resolution NHD.

definitions, flow relationships, information that links features in different NHD themes, and dataset update status.

Two route subclasses in the NHD coverage, DRAIN and RCH, are needed to extract a single-stream network coverage. The routes that build on the single or multiple arcs in route subclass DRAIN represent NHD line features that are classified as Stream/Rivers, Canal/Ditches, Pipelines, or Artificial Paths through open waters. The route subclass RCH delineates the transport and coastline reaches with similar hydrologic characteristics. Each reach has a unique reach code and may be composed of one or more routes in the DRAIN subclass. The two route subclasses are linked through the unique reach code.

The NHD data model and reach-based system supports the analysis of stream networks and watersheds with a variety of analytical GIS tools. These tools include the NHD Watershed Tool (EPA and USGS, 2004); BASINS (Better Assessment Science Integrating Point and Nonpoint Sources), WATERS (Watershed Assessment, Tracking and Environmental ResultS), and StreamStats flow prediction models (EPA and USGS, 2005); WaterNet (Olivera et al., 2006); and the TNC GIS tools (FitzHugh, 2005b) described below. The quality or "fitness for use" has emerged as a recurring issue influencing the adoption and use of the NHD with these various GIS-based analytical tools. The pre-processing routines incorporated in many of these toolsets (e.g., ArcHydro; see Maidment, 2002; NHD Watershed Tool, see EPA and USGS, 2004; TNC GIS tools, see FitzHugh, 2005b; WaterNet, see Olivera et al., 2006) signify the importance and necessity of repairing various problems resulting from the erroneous specification of the attributes of hydrographic features, such as arc-node connectivity and stream arc directionality. This paper reports our attempt to assess the extent of these problems and how much further work is needed to identify and fix the problems in the current high-resolution NHD when using these data with the TNC GIS tools to perform regional watershed assessments in southern California.

The TNC GIS tools first pre-process the NHD and generate a preliminary flow network, which can be further edited to create a "valid" flow network that meets the specifications of the tools (Fig. 2). Based on the improved stream network, the tools are then implemented to generate stream, lake, and watershed characteristics that are used with abiotic factors derived from readily available GIS data such as Digital Elevation Models (DEMs) to perform a comprehensive classification of freshwater systems. The tools may also be used to estimate the potential quality of these streams and lakes using information on the environment and human impacts such as geology, soils, land use, roads, dams, and point pollution sources (FitzHugh, 2005a). Table 1 lists input data requirements and outputs generated by the tools at different stages of tool implementation.

The extraction, interpretation, and derivation of a well-defined stream network from the original NHD datasets are key steps in performing watershed assessments using the TNC GIS tools. The TNC GIS tools offer four specific functions to assist with NHD preprocessing. The NHD Preparation Arc Macro Language (AML) script that runs in the Arc/Info environment extracts a preliminary stream network coverage in which line features, nodes, and the accompanying topology are stored. The Arc Direction and Fix Divergences tools that run inside ArcView 3.x are then used for diagnosing and fixing some or all of the problems concerned with stream network directionality, connectivity, and flow divergences. The Main_Path AML is another script that runs inside Arc/Info and is implemented when one or more alternate channels need to be deleted from the NHD stream network (Fig. 2).

SOUTHERN CALIFORNIA CASE STUDY

Our experiences show that substantial user intervention is required for editing, cleaning, and constructing the final stream network with these tools in large metropolitan areas. The Green Visions Plan study area covers five watersheds in southern California indexed by 8-digit HUC, namely Calleguas Creek, the Los Angeles River, Santa Clara River, San Gabriel River, and Santa Monica Bay (Fig. 3). The five water-

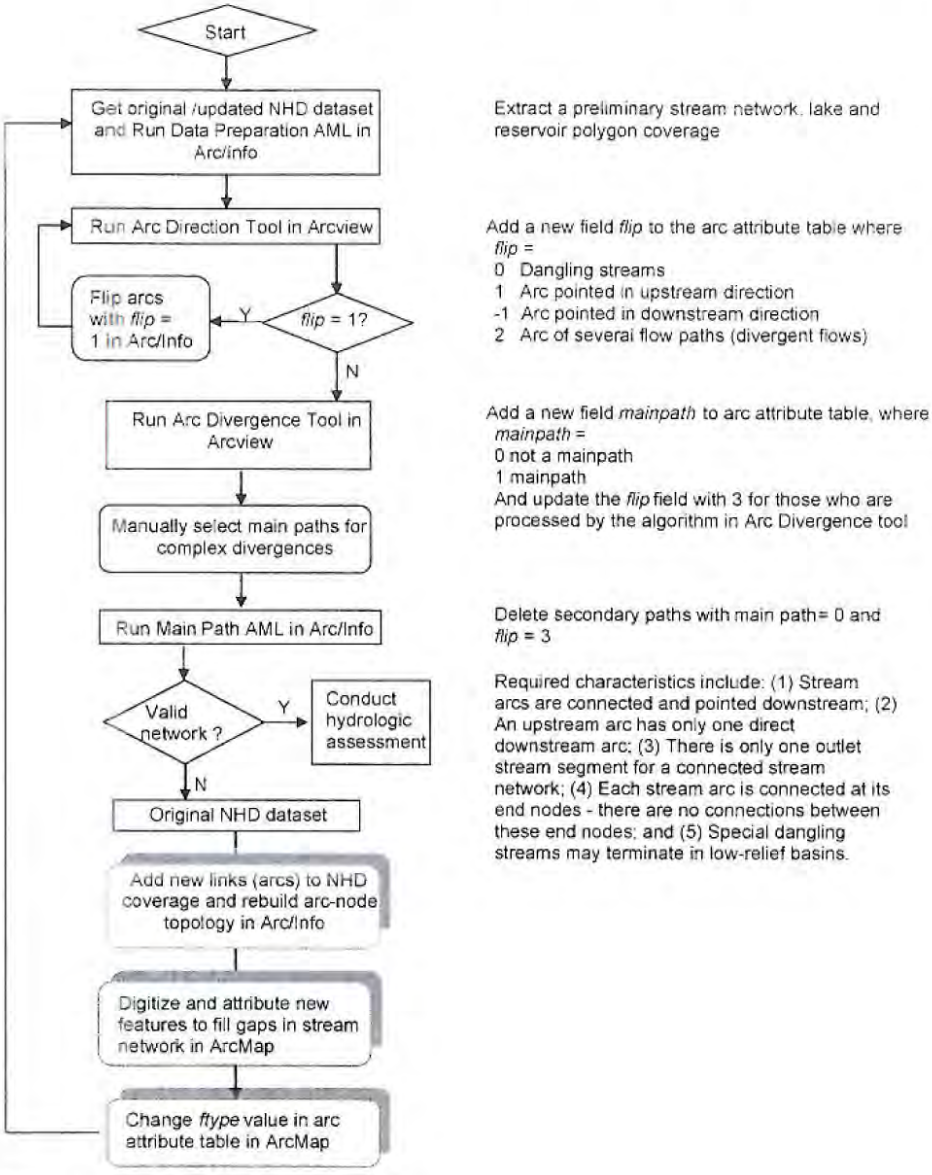

Fig. 2. Typical TNC GIS toolset workflow.

sheds drain 10,637 km^2 and the elevation varies from sea level to 3,060 m. The largest Santa Clara River watershed constitutes 39% of the study area and is the most pristine. The Calleguas Creek watershed is the smallest of the five watersheds (9% of study area) and contains several medium-sized cities as well as extensive open space (upper watershed) and agricultural areas (lower watershed). The Los Angeles River watershed is the second largest (20% of study area) and the most urbanized of the five watersheds, containing parts or all of 36 cities and 4,276,100 residents (Bhaduri et al., 2002). The San Gabriel River and Santa Monica Bay watersheds also support large

Table 1. Input Data Requirements and Attributes Generated by TNC GIS Tools

Input data requirements	Attributes (Outputs)
Stream arcs must be pointed downstream	Stream gradient and elevation
Stream arcs must be pointed downstream, and must be connected to lake arcs at nodes	Classification of stream connectivity (whether connected to a stream, lake, or coastline, or unconnected)
Lakes must exist as polygons	Lake elevation and underlying geology
Lakes must exist as polygons, and lake arcs must be connected to stream arcs at nodes	Number of stream arcs connected to each lake polygon
Stream arcs must be connected at nodes, centerlines must traverse lake and large river polygons, stream and centerline arcs must be pointed downstream, and main path must be delineated through flow divergences	Stream and lake orders and arbolate sums Stream and lake watershed area and upstream quality indicators Nested watersheds used to represent classified freshwater systems

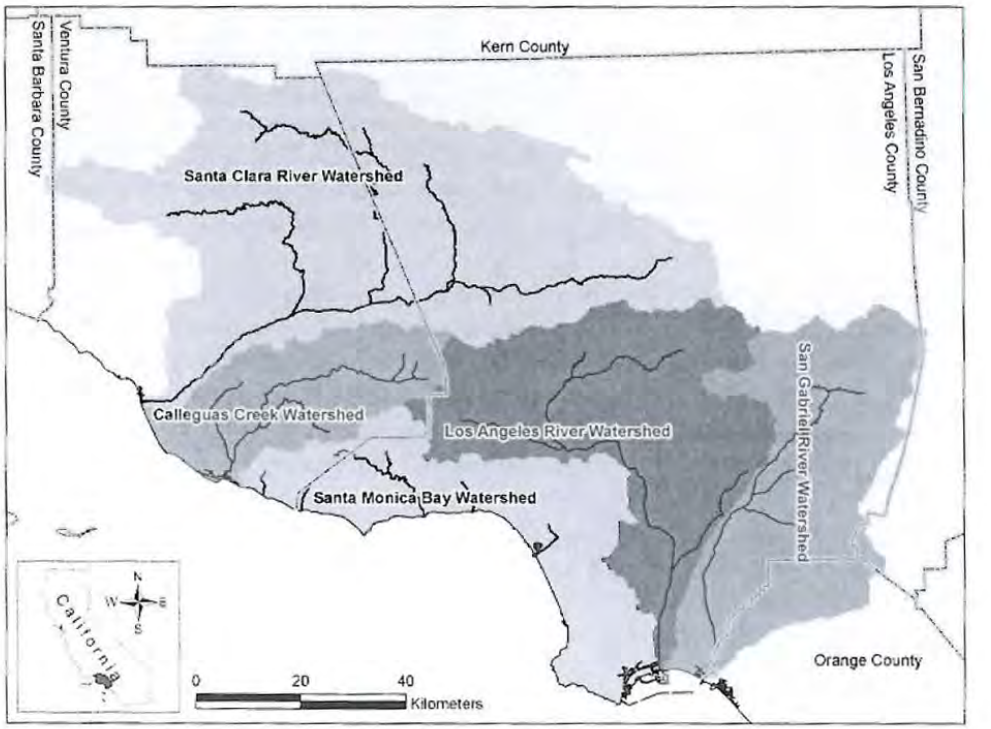

Fig. 3. Five major stream networks and watersheds in Green Vision Plan study area.

populations at lower elevations and extensive natural areas at higher elevations. The Santa Monica Bay watershed follows the coast and consists of numerous small streams or rivers that empty into the Pacific Ocean (Fig. 3).

Table 2. Basic Watershed Characteristics (in km) Following First Run through TNC GIS Tools

	Watershed					
Characteristics	Calleguas	Los Angeles	San Gabriel	Santa Clara	Santa Monica	Totals
Stream/River	1,222.7	2,060.5	1,887.8	6,311.1	1,344.7	12,826.8
Lake Path	27.1	30.4	33.1	88.8	25.1	204.5
Pipeline/ Canal/Ditch	0	0	0	0	0	0
Totals	1249.8	2090.9	1920.9	6399.9	1369.8	13,031.3

The TNC GIS tools were run multiple times for each of these southern California watersheds. For the first run, the original high-resolution NHD for the study area was downloaded from the NHD website [ftp://nhdftp.usgs.gov/FOD_Data/High/arc/] and processed using the NHD Preparation AML in Arc/Info and Arc Direction and Fix Divergences tools in ArcView. This run yielded a preliminary network coverage that was extracted from the network theme with the Pipeline or Canal/Ditch features deleted. The resulting network layer contains the feature types of Stream/River, Lake Path, and Coastline that were reclassified from the original NHD coverage. Table 2 summarizes the length of the Stream/River and Lake Path feature types in the five watersheds following the first run through the TNC GIS tools.

The resulting stream network did not satisfy the valid stream network requirements for watershed assessments with the GIS TNC tools in our study area (Fig. 2). In addition to the problems highlighted in the tutorial that accompanies the TNC GIS tools, there were more problems and errors found in the study area. Features with erroneous flow directions, divergences, and no downstream connections were also distinguished from the remainder of the stream network during this run (see Tables 3 and 4 below for additional details).

Problems

The initial TNC GIS run identified 4,234 problems or errors in the high resolution NHD (Tables 3 and 4). The majority of these problems (79.5%) were dangling or missing stream segments (reaches). Some of these dangling streams ended in special areas (flat coastal areas and low relief sub-basins), although most (3,048 of 3,366) were scattered throughout the study area (Table 4). This last result may reflect the Mediterranean climate and/or urban character of the study area, as will be discussed later.

The second most common class of problems included simple and complex divergences. Simple flow divergences occur where a channel splits into two branches along part of its course, whereas complex divergences occur when the channel splits into three or more branches (Fig. 4). It was fortunate that the simple case (which can

Table 3. Major Problems Encountered in Green Visions Plan Study Area

Type of problem		Detailed description
Dangling and/or missing segments (reaches)		Dangling streams ending in flat coastal areas
		Dangling streams ending in low-relief basins
		Dangling streams ending in other parts of landscape
Simple/complex divergences		Streams diverge and meet again
Attribute errors	Attribute designation errors	Stream reaches assigned wrong FTYPE code
	Flow direction errors	Flow directions pointed upstream
	Unnecessary removal of reaches	Removal of pipelines, canals and ditches by TNC GIS tools caused new network connectivity problems
Duplicate stream segments (reaches)		A unique stretch of stream/river was represented twice by parallel paths

be identified and repaired automatically with the TNC GIS tools) occurred many more times than the complex divergences (Table 4).

The third class of problems consisted of attribute errors. Some of these problems occurred because some stream reaches had been assigned the wrong FTYPE codes, some occurred because stream reaches had been assigned flow directions that pointed upstream, and others occurred because they represented pipeline, canal, and ditch reaches that connected streams and rivers but were automatically removed from the "final" NHD stream network when the TNC GIS tools were run. This third problem accounted for nearly 70% of the problems in this class (Table 4).

The final class of problem took the form of duplicate stream segments (reaches). These problems were relatively rare and occurred in just two of the five watersheds where the NHD recorded two parallel streams when only one was present (Fig. 5).

Solutions

The three grey-shaded boxes near the bottom of Figure 2 represent the tasks that were performed to repair these problems. Following the specifications of the TNC GIS tools, the changes to the network—including editing, adding, removing features, and building arc-node topology—were made within the existing NHD coverage to ensure the proper performance of the NHD Preparation and follow-up watershed assessment tools. Five new datasets were acquired and used to support the repairs as follows. USGS 10 m DEMs were downloaded from the GIS Data Depot [http://data.geocomm.com/]. A line coverage containing urban storm drains owned and operated by Los Angeles County was obtained from the San Gabriel and lower Los

Table 4. Frequency of Different Types of Problems and Errors by Watershed Following First Run through TNC GIS Tools[a]

Problem		Calleguas		Los Angeles		San Gabriel		Santa Clara		Santa Monica Bay	
		#	%	#	%	#	%	#	%	#	%
Dangling and/or missing segments (reaches)	Ending in flat coastal areas	35	1.5	0	0.0	0	0.0	0 0	0.0	203	14.5
	Ending in low-relief basins	0	0.0	18	0.9	33	1.8	0	0.0	29	0.8
	Ending in other parts of landscape	341	25.6	576	21.3	1224	66.1	721	13.5	186	11.4
Flow divergences	Simple divergences	27	0.8	74	0.5	25	0.8	419	1.9	14	0.2
	Complex divergences	0	0.0	0	0.0	0	0.0	35	0.1	0	0.0
Attribute errors	Attribute designation errors	30	0.7	7	0.2	4	0.1	0	0.0	0	0.0
	Flow direction errors	6	0.2	10	1.1	14	0.4	5	0.0	0	0.0
	Unnecessary removal of reaches	29	1.5	57	3.1	84	3.6	7	0.1	2	0.3
Duplicate stream segments (reaches)		0	0.0	3	1.2	15	0.4	0	0.0	0	0.0
Totals		468	30.3	745	28.3	1399	73.2	1187	15.6	434	27.2

[a]The first column within each watershed indicates the number of network routes (#) that were affected, and the second column shows the length of these routes as a percentage (%) of the total stream length in each watershed.

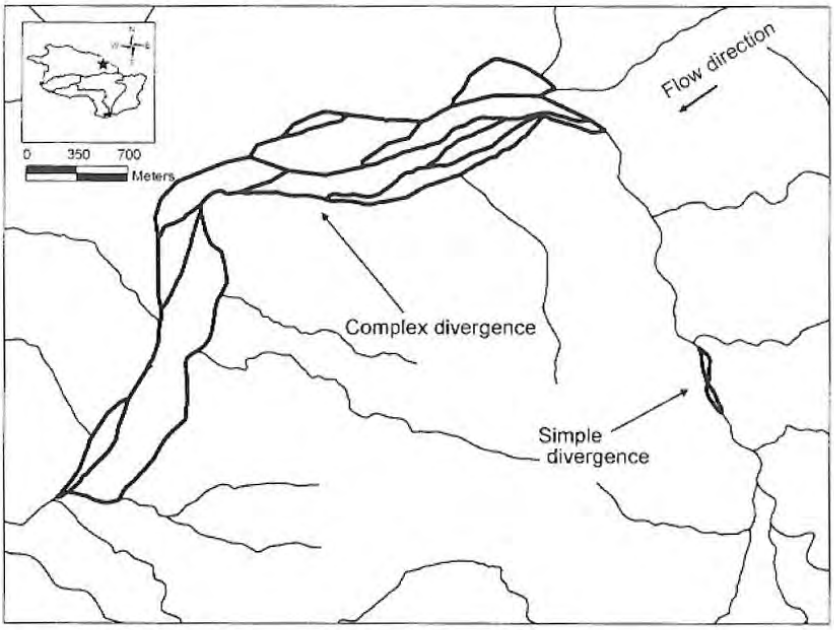

Fig. 4. Part of high-resolution NHD stream network containing complex and simple divergences.

Fig. 5. Portion of lower Los Angeles River represented in high-resolution NHD as both a Stream/River feature (Reach A) and an Artificial Path feature (Reach B).

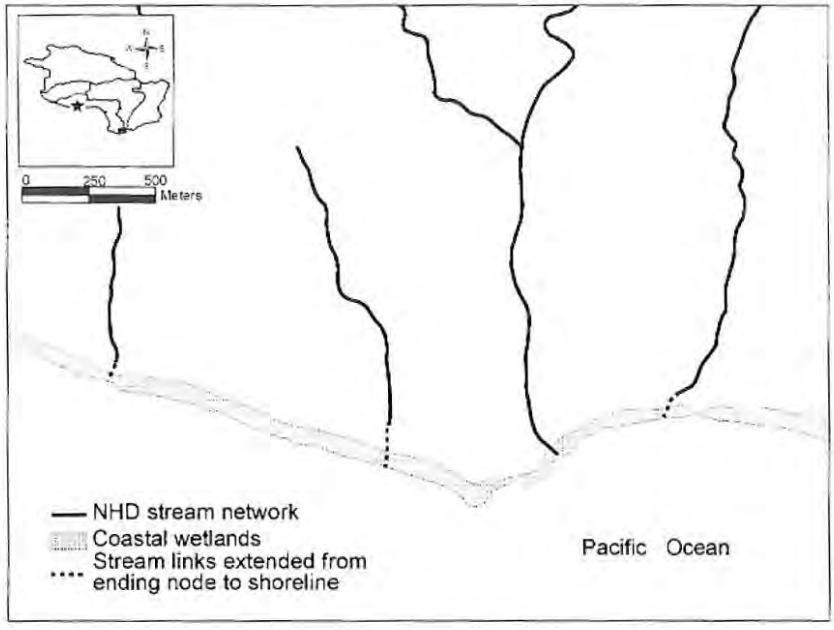

Fig. 6. Dangling streams in flat areas near the Pacific Ocean were extended and snapped to shoreline to extend connectivity of the stream network.

Angeles rivers and Mountain Conservancy (RMC). Raster images of scanned USGS 1:24,000-scale topographic maps, also known as Digital Raster Graphics (DRG) files, were downloaded from the California Spatial Information Library [http://casil.ucdavis.edu/casil/gis.ca.gov/drg/]. High-resolution (1 m) aerial photography from 1999 and 2000 was acquired from the RMC. National Wetlands Inventory (NWI) datasets were downloaded [http://www.fws.gov/data/statdata/cadata.html] and used to identify coastal estuaries. All these datasets were re-projected to Universal Transverse Mercator Projection (UTM), zone 11, using the North American Datum of 1983, to match the high-resolution NHD. The following sections describe the methods and data sources utilized to fix the four classes of problems (errors) summarized in the previous section.

Dangling Streams. Three classes of dangling streams were identified across the five watersheds (Table 4) and the bulk of them (91.5% when measured in terms of the length of the affected stream reaches) were eliminated using the available reference flow networks and related terrain and land use/land cover data. Streams ending near coastal features were assumed to discharge to the adjacent ocean and therefore the NHD stream network was extended to fill these gaps. The National Wetlands Inventory dataset was used to check the connectivity of coastal streams to marshes, wetlands, tidal inlets, and estuaries along the Pacific Ocean coastline. The end nodes of 34 dangling coastal streams were selected, edited, and snapped to the shoreline following the shortest path in ArcEdit (Fig. 6). This rebuilt the connectivity for 238 network routes with a total length of 217.5 km (Table 4).

The second class of dangling streams, those that drained to low-relief sub-basins, was assumed to occur naturally. There were 80 such reaches with a total length of

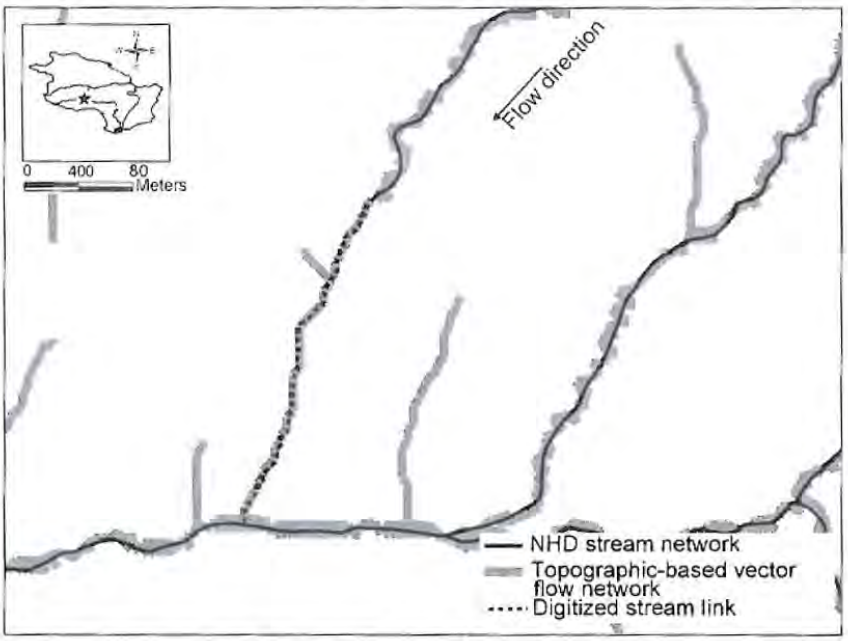

Fig. 7. Gaps in high-resolution NHD stream network that were filled using part of topographic-based vector flow network.

64.8 km (Table 4) that terminated in retarding basins, detention basins, percolation basins, lakes, and reservoirs and lacked connections to the remainder of the network. No action was taken to extend these streams to the coast or to link these streams to other parts of the stream network. These were left unaltered and were not counted in the subsequent stream network leveling and tracing. They consequently form isolated drainage basins nested within larger drainage basins.

Three sets of stream networks that were generated from or delineated in other geospatial datasets were used to repair the remainder of the dangling streams. Topographic-based stream networks computed from USGS 10 m DEMs were used to fill many of the gaps in the network. The ANUDEM 5.1 software developed by Hutchinson (1989, 2004) was applied to produce an accurate DEM with hydrologic properties from point elevations, stream lines, and lake polygons. The program also produced a flow direction grid from the output DEM using the D8 algorithm developed by O'Callaghan and Mark (1984). It distinguished each DEM pixel as either a stream or a non-stream pixel. The NHD stream network was overlaid on a topographic-based stream network and the missing links between dangling streams and the remainder of the network were filled by following the topographically derived flow paths (Fig. 7). A total of 634 new routes (mean length = 0.79 km) was added using this approach (Table 5).

Urban storm drains were used to fill in some of the gaps in urban landscapes, where natural flow paths have been severely altered or replaced by storm drains and the topographic-based stream network did not accurately represent the existing surface water drainage in low relief areas. The storm drains were particularly important because they offered information to determine flow paths at locations where natural

Table 5. Numbers of Added Routes (#) and Percentage Share of Length of Digitized Links of the Total Stream Length in Each Watershed (%)

Reference datasets	Calleguas		Los Angeles		San Gabriel		Santa Clara		Santa Monica Bay	
	#	%	#	%	#	%	#	%	#	%
Topographic-based vector network	89	4.1	64	3.0	64	2.5	373	4.4	44	4.3
Urban storm drains	0	0.0	73	6.9	58	2.7	0	0.0	37	5.9
Assorted maps and field observations	31	2.1	27	0.8	19	1.0	0	0.0	10	0.5
Total	120	6.2	164	10.7	141	6.2	373	4.4	91	10.6

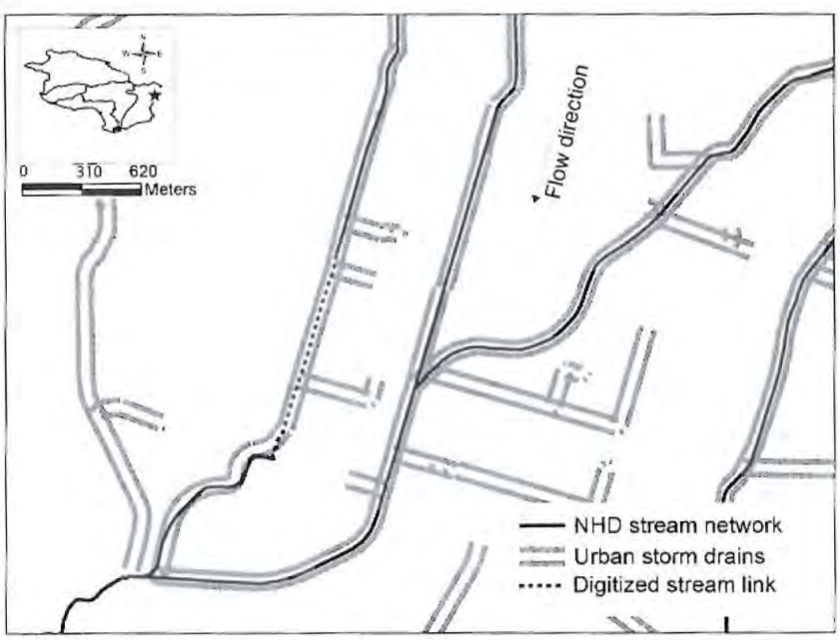

Fig. 8. Gaps in high-resolution NHD stream network that were filled using part of urban storm drain dataset.

stream flow was rerouted underground (Fig. 8). A total of 168 new routes with a length of 277 km were added to the Los Angeles River, San Gabriel River, and Santa Monica Bay watersheds (i.e., the most urbanized areas) using this approach (Table 5).

The remainder of the gaps in the stream network in urban areas was filled using evidence compiled from topographic maps, aerial photographs, and field inspections. Contour lines on 1:24,000-scale topographic maps assisted with the definition of potential flow paths that were not delineated in the topographic-based network. Lines of steepest descent were determined on the contour maps and used to extend and

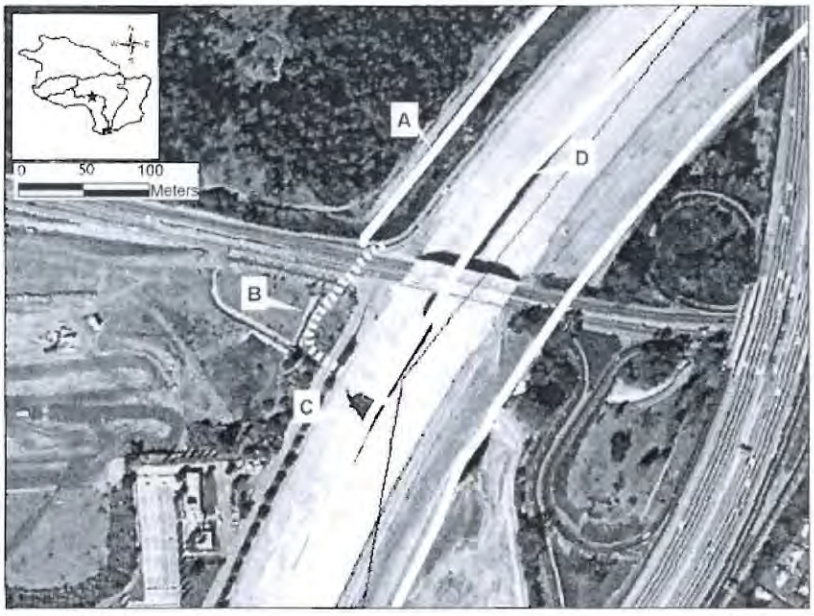

Fig. 9. Gaps in NHD stream network that were filled using ground information discerned from aerial photographs. Label A denotes a dangling stream in the NHD network. The dashed line labeled B points to a missing link that was added with heads-up digitizing. The point labeled C indicates the water footprint on the image, and D shows the confluence of the main and tributary streams. The digitized link was offset from the underlying water path in order to see it clearly on the image.

connect dangling streams to downstream network reaches. Aerial photographs provided additional information about open water features, riparian vegetation, and vegetated corridors, and showed in particular that: (1) the drainage pattern in urban areas tended to follow major roads, riparian and vegetated corridors; and (2) the streams usually emerged above ground when they drained to the main tributaries in suburban areas. The water marks discerned from the aerial photograph in Figure 9, for example, clearly demonstrate the connectivity of the stream to the main channel after the stream passed under a bridge. A total of 81 new routes with a length of almost 71 km were added to the NHD stream network using this approach (Table 5).

Seven tasks were performed to add new features to the network and rebuild the arc-node topology of the updated coverage in Arc/Info:

1. Heads-up digitizing was performed to fill the gaps in the network in a newly created shapefile layer by referring to the appropriate reference flow network data.

2. Newly added features were thereafter properly routed to ensure the correct arc-node topology was built at the data preparation stage. The shapefiles containing digitized line features were also converted to two separate route coverages.

3. All items in the route subclasses of these two new coverages were modified to match the items in the DRAIN and RCH route subclasses in the NHD coverage.

4. New features inherited their attributes from upstream reaches in most cases with the exception of the FTYPE and subclass-ID attributes.

5. The FTYPE codes of all newly digitized features were defined as Connector. The subclass ID assigned to each new feature was a number beyond the range of the subclass IDs in the DRAIN and RCH themes. Downstream attributes were utilized in some cases because the attributes of the added paths were more like the downstream reach than the upstream reach. New features that connected lake paths, artificial paths, or manmade flow paths to downstream natural flow paths were treated this way, for example.

6. Two new route subclasses were added to the original NHD coverage route subclasses DRAIN and RCH, respectively, in the ArcEdit module of Arc/Info.

7. The Build command was used to add topology to the updated NHD coverage in Arc/Info.

The additional routes (links) added along the coast and with the assistance of the three reference datasets in other parts of the study area eliminated 3,242 of the 3,366 dangling streams. Most of the remaining dangling streams (80 out of 124 routes) ended in low-relief basins, and those still not accounted for (44 routes with a total length of 21.4 km) were deleted from the network given their small number, short lengths, and because no connectivity information could be retrieved from any readily available reference data source in these instances. The total length of deleted dangling streams accounted for less than 0.3% (by length) of the total stream network in the study area.

Flow Divergences. We utilized the Arc Direction tool in the TNC GIS tool sets to check for the presence of flow divergences as shown in Figure 4. A new attribute (FLIP) was added to the stream network coverage to indicate whether flow divergences were present (Fig. 2). The Fix Divergences tool was then used to distinguish simple divergences and complex divergences (by updating the value of the FLIP attribute). The flow divergence problems were fixed with the TNC GIS tools or through user intervention after the Arc Direction and Fix Divergences tools were run. The 594 simple divergences located with the aforementioned tools were easily repaired by running the Main_Path AML in Arc/Info (Fig. 2). This tool selected the shortest paths as the main paths, deleted the secondary paths, and rebuilt the arc-node topology in ArcEdit in these instances. The 35 complex divergences were all located in the Santa Clara River watershed and were modified by manually selecting the arc representing the main flow path and deleting the arcs representing the secondary flow paths in ArcEdit. The determination of the main flow paths for a braided river is difficult in reality because the main paths vary in space and time. The principle we used was to keep the shortest path, which is also the path with the steepest slope, based on the premise that this channel was the most important path for hydrologic modeling.

Attribute Errors. The Arc Direction tool run from ArcView was also used to check stream arc direction (Fig. 2). A total of 24 flow direction errors was diagnosed with the Arc Direction tool in four of the five major watersheds (the Santa Monica Bay watershed had no flow direction errors). The arcs with erroneous flow directions were relatively short (27.0 km total length). Another 11 arcs (6.2 km total length) were detected through the manual checking of dangling streams. The flow directions for all the routes with reversed directions were selected and flipped in ArcEdit using the FLIP command.

A total of 41 reaches were found to have attribute designation errors when features with inconsistent geometry were examined. In the high-resolution NHD, natural

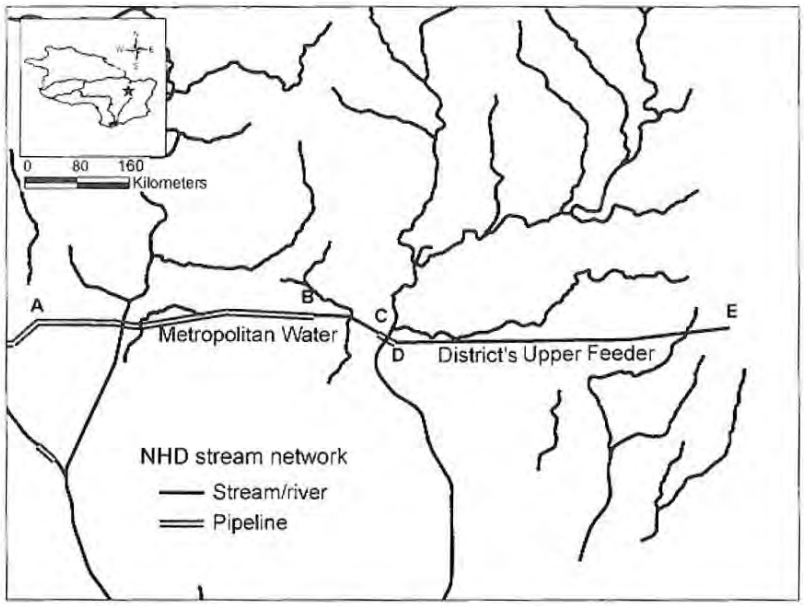

Fig. 10. Attribute designation errors in the NHD network. The FTYPE codes for the stream segments labeled AB and CD were set to Pipeline, and for BC and DE to Stream/River. However, all four segments from A to E represented part of the Metropolitan Water District's Upper Feeder.

streams have a curvilinear form and seldom show up as straight-line segments in aerial photographs or field inspections. This observation allowed detection of some streams that were labeled incorrectly as pipelines, canals, or ditches, and located mostly in the Calleguas Creek watershed. In some other situations, pipelines, canals, and ditches were designated incorrectly as streams, such as that shown in Figure 10. The FTYPE codes for those features were changed according to the information interpreted from the topographic map. These changes meant that 17.0 km of stream reaches were switched from Streams/Rivers to Pipelines or Canals/Ditches and subsequently deleted when the NHD Preparation AML was run a second time.

The third set of attribute problems occurred because the NHD Preparation AML automatically deleted arcs with FTYPE codes set to Pipelines and Canals/Ditches when many of these arcs were required to sustain the connectivity of the overall network. Nearly 200 Pipeline and Canal/Ditch routes with a total length of 158.8 km were deleted from the network by the execution of the NHD Preparation AML. Figure 11 shows two stream links (A and B) that were removed at this step. This action generated additional gaps in the network besides those present in the original NHD network subclass (as discussed earlier). The Pipelines and Canals/Ditches that contributed to the connectivity of the network were therefore selected manually and kept as part of the network by changing the FTYPE codes to Connector.

Duplicate Stream Segments. Eighteen duplicate streams like that shown in Figure 5 were identified (Table 4). These occurred in highly urbanized areas along the lower Los Angeles River and the Rio Hondo Channel and the main channel of the San

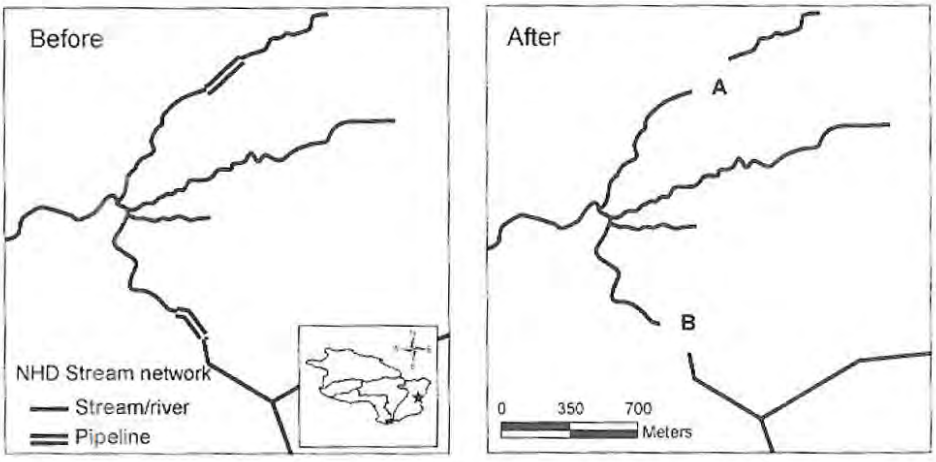

Fig. 11. Unnecessary removals of network reaches. The gaps labeled A and B resulted when execution of the TNC GIS Data Preparation tool removed the FTYPE code of Pipeline from the NHD network

Gabriel River. Their absence on the ground was verified from aerial photographs and topographic maps and the attribute values for FTYPE were changed to Canal/Ditch features for the least important (i.e., least connected) of each pair of parallel stream segments. This approach was effective because the features with FTYPE codes of Canal/Ditch were deleted automatically when the NHD Preparation AML was run a second time (Fig. 2).

Final Run

The TNC GIS tools were run a second time using the improved NHD dataset. The remaining attribute errors were corrected and the dangling streams were removed using the Arc/Info AMLs and ArcView tools (Fig. 2). The differences in basic stream network statistics between first and final runs through the GIS TNC tools are summarized in Table 6, and show how the stream network in five watersheds was modified from start to finish after digitized links were added to fill the gaps in the original stream network and the other types of modifications were applied. The lengths of Stream/River were increased by an average of 95.7 km for each watershed. Very few changes were made to Lake Path but for the exception of removal of isolated lake paths from the network. A total of 143.4 km of pipelines, canals, and ditches were retained in the stream network in the last run of the TNC GIS tools, instead of their total removal as occurred during the first run of the tools. Overall, the total length of the revised stream network increased by as much as 11.3% in the Los Angeles River watershed and as little as 1.2% in the Santa Clara River watershed (Table 6).

Comparison of Watershed Attributes Generated with Original and Revised NHD

Two sets of stream networks, one generated from the original NHD and the other from the enhanced NHD, were implemented with watershed boundary delineation and flow accumulation area tools to compare the flow accumulation results at a series

Table 6. Differences in Basic Stream Network Statistics between the First and Final Runs through the TNC GIS Tools

	Watershed					
Characteristics (km)	Calleguas	Los Angeles	San Gabriel	Santa Clara	Santa Monica Bay	Totals
Stream/River	68.6	183.0	78.2	71.6	77.2	478.6
Lake Path	–0.7	0.9	–1.3	1.4	–0.3	0.0
Pipeline/Canal/Ditch	20.3	48.0	67.6	4.0	3.5	143.4
Totals	88.2	231.9	144.5	77.0	80.4	622.0

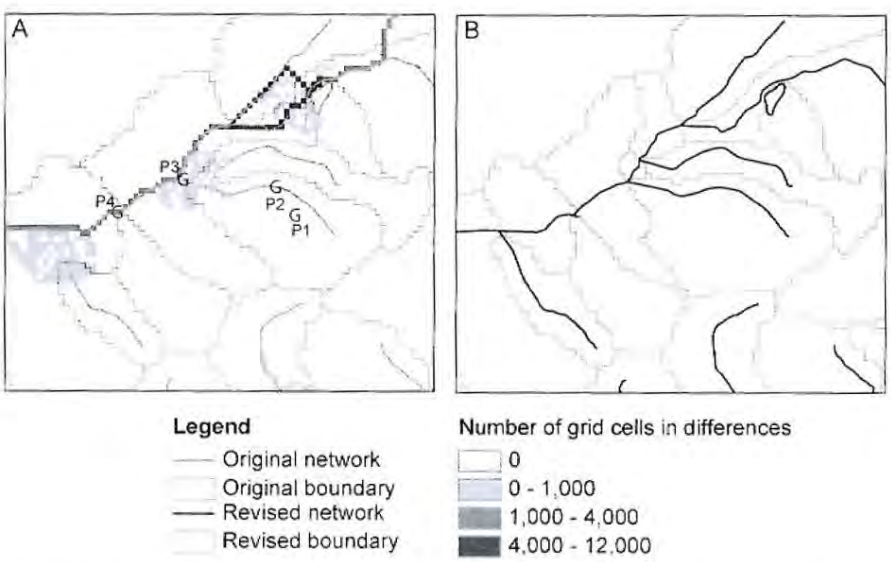

Fig. 12. A. Revised NHD network and its generated watershed boundaries. Points P1 to P4 are four grid cells sampled at different hillslope positions that indicate varying hydrologic processes. The magnitude of the differences in flow accumulation areas caused by using the original NHD and revised NHD are shaded in greyscale. B. Original NHD network and its generated watershed boundaries.

of downslope points. The upslope contributing area (i.e., flow accumulation) is a critical parameter for determining runoff, water quality, and flood potential within a given watershed. The watershed boundary was delineated using the catchment delineation AML in the TNC GIS tool kit. The flow accumulation in each grid cell was calculated inside ArcGIS 9.1 using the flow accumulation function. The differences between the two sets of watershed boundaries and flow accumulation estimates are shown in Figures 12A and 12B. These maps show small areas in the San Gabriel

Table 7. Numbers and Mean Watershed Areas Delineated with the TNC GIS Tools and Original/Revised NHD[a]

Strahler stream orders	# of watersheds		Total watershed area, km^2		Mean watershed area, km^2	
	Original NHD	Revised NHD	Original NHD	Revised NHD	Original NHD	Revised NHD
0	1819	421	1036.8	151.6	0.57	0.36
1	308	966	268.0	773.0	0.87	0.80
2	144	498	217.4	259.0	1.51	0.52
3	77	352	114.7	260.5	1.49	0.74
4	61	159	103.1	217.8	1.69	1.37
5	22	44	59.2	53.2	2.69	1.21
6	–	39	–	75.7	–	1.94

[a]Zero-order streams indicate streams that were not processed by the watershed delineation tool because of lack of connections to main tributaries that empty into the ocean.

River watershed. The watershed boundaries generated with the two NHD coverages were similar in the mountainous areas with high relief. The discrepancies grew larger as the terrain became flat, urban settlements appeared, and significant alteration of the surface drainage systems occurred.

The number of watersheds and average watershed size are summarized in Table 7 by Strahler stream order. The total drainage areas estimated with the two NHD coverages remained the same (i.e., 1,799 vs. 1,791 km^2), as expected. However, the area of non-processed watersheds (or zero-order watersheds) was reduced from 58% to 8.5% of the total watershed area when the revised NHD was used in place of the original NHD. The main channel was classified as a 6th-order Strahler stream using the revised NHD and the number of watersheds in the 1st- through 6th-order categories increased as a consequence of the re-ranking of watersheds that were not previously processed. The mean watershed sizes for Strahler Stream orders 1 through 6 were reduced after the boundary was redrawn as well.

Watershed attributes like flow accumulation area are likely to receive different values when watershed boundaries are varied. Four sample points labeled P1 through P4 were selected and their flow accumulation values were compared (Fig. 12). These points are taken from different hillslope positions that represent various underlying hydrologic processes: the first is located inside a headwater catchment (P1); the second in a first-order stream channel (P2); the third near the outlet of a dangling stream where the watershed boundary was moved using one or other of the NHD covereges (P3); and the final point is located in a downstream channel to which new tributaries were added when the revised NHD was used in place of the original NHD (P4). Identical flow accumulation values were calculated with the two NHD coverages at points P1 and P2. Larger flow accumulation values were calculated at P3 (12.5%) and P4 (17 times) using the revised NHD coverage. These results coupled with the differences in

shading in Figure 12 show some of the consequences of using the revised NHD (with its better network) in place of the original NHD.

DISCUSSION AND CONCLUSIONS

The four classes of problems—dangling streams, attribute errors, flow divergences, and duplicate streams—occurred throughout the five watersheds comprising the study area. Some of the problems were more prevalent in some parts of the landscape than in others. The problems with dangling streams were ubiquitous, but those occurring in natural and high-relief lands were more likely to be repaired by referring to the topographic-based flow network. The repair of those occurring in urban areas was linked to the availability and quality of urban storm drain datasets. The dangling streams in the coastal area were assumed to discharge directly into the ocean, which greatly simplified the process of repairing the stream network in these instances. Most of the problems with flow divergences, on the other hand, occurred in the relatively pristine Santa Clara River watershed, where braided streams are common. The attribute errors, including reversed flow directions, designation errors, and the unnecessary (unwanted) removal of pipelines, canals, and ditches were scattered across the study area but were more prevalent in urban areas due to the complexity of the drainage systems and difficulties encountered identifying surface water features. Duplicate stream segment problems were relatively rare in these watersheds and were easily fixed by removing the least important paths in the duplicate pairs.

The final stream network was a single-line natural flow network with correct arc-node topology and no dangling streams, with the exception of streams terminating in the low relief basins, and few, if any, attribute designation problems in the watershed. The case study we have conducted suggests that the revised NHD data can be used with the TNC GIS tools for regional watershed assessments in large metropolitan region like southern California. Our study showed that the best results would be obtained after checking for the presence of geometric as well as attributes errors. The connectivity should be re-established in the database with reference to ground data once these errors are fixed to ensure the appropriate stream classification and relevant watershed attribute generation. Our study showed that the pipelines, aqueducts, and canals that sustain the hydrologic connectivity should be identified and retained in the stream network to ensure the integrity of the watershed units in representing hydrologic features on the ground. This is likely to be true of any landscape in which the stream network and accompanying water features (wetlands, ponds, etc.) have been extensively modified to accommodate the built environment.

From a GIS perspective, all modifications should be made on the existing NHD route subclasses of DRAIN and RCH and the accompanying feature class attribute tables to maintain the NHD spatial organization schema. Features should be added to or removed from both the DRAIN and RCH subclasses to ensure the integrity of the NHD coverage and subsequent data preparation using the TNC GIS tools. In cases where multiple source data sets are available for reconstructing the flow network, priorities should be determined based on the quality of source data. We assumed that: (1) the aerial photography-interpreted stream network data were superior to all other sources; and (2) the storm drain data were superior to the topographic-based network data and topographic maps in urban areas. There are some issues that need to be dealt

with in the NHD and TNC GIS tools application. Features that were removed from the NHD manually or using the TNC GIS tools should somehow be represented in the stream network, if those features exist on the ground. Our approach helped determine the main and secondary paths and facilitated the preparation of a satisfactory network to support watershed assessments across large regions. However, divergences are common in braided channel systems and we therefore need a better way to represent and analyze flow divergences in stream flow networks. Some of the pipelines, canals, and ditches that were deleted from the system with the TNC GIS tools may have caused information loss if these features actually exist on the ground. It was also noticed that some valuable attributes attached to the reaches were not inherited correctly from the original NHD to the corresponding revised stream network reaches, such as the designation of stream types and codes. The rebuilding of the arc-node topology and removal of pseudo-nodes on the arcs in Arc/Info caused uncertain designation of stream type codes to the merged or split arcs. Further studies are required to detect the loss and modification of the NHD attributes during these operations and to ensure that the attributes are appropriately passed to the newly generated features in the revised stream network. A bottom-up approach, in which local experts and data sources are linked with NHD database designers and engineers as well as applicable builders, should be adopted to refine the NHD network and verify the flow delivery and stream attributes.

REFERENCES

Bhaduri, B., Bright, E., and P. Coleman, 2002, "LandScan USA: A High Resolution Population Distribution Model," *Computing Science and Statistics*, 34 [/I2002Proceedings/BhaduriBudhendra/BhaduriBudhendra.abstract.pdf], accessed Mar 20, 2006.

Dewald, T. G. and K. S. Roth, 1998, "The National Hydrography Dataset—Integrating the USEPA Reach File and USGS DLG," in *Proceedings: 1998 International ESRI User Conference*, Redlands, CA: Environmental Systems Research Institute.

EPA (U.S. Environmental Protection Agency) and USGS (U.S. Geological Survey), 2000, "The National Hydrography Dataset, Concepts and Contents" [http://nhd.usgs.gov/chapter1/index.html], accessed February 12, 2005.

EPA and USGS, 2004, "NHD Watershed Tool—Instructions for Preprocessing Supporting Data Layers" [http://nhd.usgs.gov/watershed/watershed_tool_inst_TOC.html], accessed Feb 12, 2005.

EPA and USGS, 2005, "NHD Applications" [http://nhd.usgs.gov/applications.html], accessed February 12, 2005.

FitzHugh, T. W., 2005a, "GIS Tools for Freshwater Biodiversity Conservation Planning," *Transactions in GIS*, 9(2):247–263.

FitzHugh, T. W., 2005b, "GIS Tools for Stream and Lake Classification and Watershed Analysis," Arlington, VA: The Nature Conservancy [http://conserveonline.org/workspaces/gistoolsfreshwater], accessed May 10, 2005.

Gadd, S., 2002, "Partnership for Development of High Resolution National Hydrography Dataset on Forest Service Lands," US Department of Agriculture, US Forest Service, National Resource Information System Water Module, Corvallis,

OR [www.fs.fed.us/emc/nris/water/nhd_lib/gen_overview.htm], accessed June 5, 2005.

Groves, C., Valutis, L. L., Vosick, D., Neely, B., Wheaton, K., Touval, J., and B. Runnels, 2000, "Designing a Geography of Hope," Arlington, VA: The Nature Conservancy, 85 [http://conserveonline.org/docs/2000/11/GOH2-v1.pdf], accessed June 5, 2005.

Hargrove, B., Tear, T., and L. Landon, 2002, "Theoretical Constructs for Large-Scale Restoration: An Accompaniment to Geography of Hope Update. When and Where to Consider Restoration in Ecoregional Planning," Arlington, VA: The Nature Conservancy [http://conserveonline.org/docs/2002/09/GOH_Update_sept02.pdf], accessed June 5, 2005.

Hutchinson, M. F., 1989, "A New Method for Gridding Elevation and Streamline Data with Automatic Removal of Pits," *Journal of Hydrology*, 106:211–232.

Hutchinson, M. F., 2004, *ANUDEM Version 5.1 User Guide*, Canberra, Australian: Centre for Resource and Environmental Studies, Australian National University.

Maidment, D. R., 2002, *Arc Hydro: GIS for Water Resources*, Redlands, CA: ESRI Press, 187 p.

O'Callaghan, J. F., and D. M. Mark, 1984, "The Extraction of Drainage Networks from Digital Elevation Data," *Computer Vision, Graphics, and Image Processing*, 28:323–344.

Olivera, F., Koka, S., and J. Nelson, 2006, "WaterNet: A GIS Application for the Analysis of Hydrologic Networks Using Vector Spatial Data," *Transaction in GIS*, 2006, 10(3):355–375.

Roni, P., Timothy, J. B., Robert, E. B., Frank, E. L., Michael, M. P., and R. P. George, 2002, "A Review of Stream Restoration Techniques and a Hierarchical Strategy for Prioritizing Restoration in Pacific Northwest Watersheds," *North American Journal of Fisheries Management*, 22:1–20.

Wittner, E., and L. McKee, 2002, *Mapping Urbanized and Rural Drainages in the Bay Area: A Tool for Improved Management of Storm Water Contaminants Derived from Small Tributaries*, Oakland, CA: San Francisco Estuary Institute, 44 p.