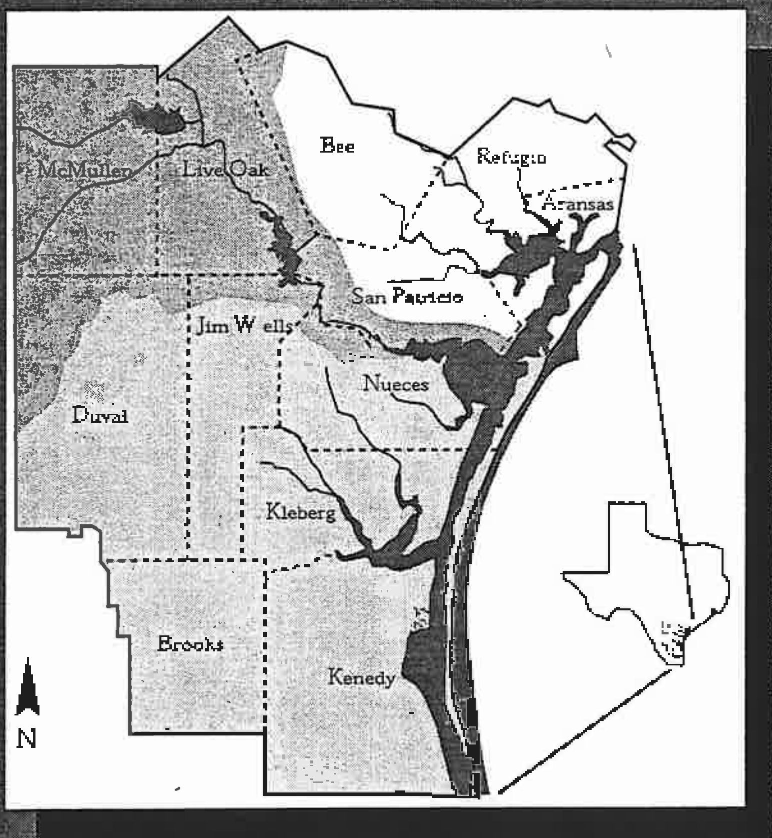
Ambient Water, Sediment, and Tissue Quality of the Corpus Christi Bay Study Area: Present Status and Historical Trends



Corpus Christi Bay National Estuary Program

CCBNEP-23 • November 1997

Ambient Water, Sediment, and Tissue Quality of the Corpus Christi Bay Study Area

Present Status and Historical Trends

Principal Investigators:

George H. Ward, Ph.D.
Neal E. Armstrong, Ph.D., P.E.. DEE

Center for Research in Water Resources
The University of Texas at Austin

Publication CCBNEP - 23
November 1997

Policy Committee

Commissioner John Baker
Policy Committee Chair
Texas Natural Resource Conservation Commission

Mr. Jerry Clifford
Policy Committee Vice-Chair
Acting Regional Administrator, EPA Region 6

The Honorable Vilma Luna
State Representative

Commissioner Ray Clymer
Texas Parks and Wildlife Department

The Honorable Carlos Truan
Texas Senator

Commissioner Garry Mauro
Texas General Land Office

The Honorable Josephine Miller
County Judge, San Patricio County

Commissioner Noe Fernandez
Texas Water Development Board

The Honorable Loyd Neal
Mayor, City of Corpus Christi

Mr. Ray Allen
Coastal Citizen

Mr. Bernard Paulson
Coastal Citizen

Management Committee

Mr. Ray Allen, Co-Chair
Mr. Bob Wallace, Co-Chair

Local Governments Advisory Committee

Commissioner Gordon Porter, Chair
Mr. James Dodson, Vice-Chair

Scientific/Technical Advisory Committee

Dr. Terry Whitledge, Chair
Dr. Wes Tunnell, Vice-Chair

Citizens Advisory Committee

Mr. Robert Corrigan, Chair
Mr. Jay Reining, Vice-Chair

Financial Planning Advisory Committee

Dr. Joe Moseley, Chair

Program Director

Mr. Richard Volk

Natural Resources Center, Suite 3300, TAMU-CC • 6300 Ocean Drive • Corpus Christi, TX 78412
512/980-3420 • FAX 512/980-3437 • CCBNEP home page: //www.sci.tamucc.edu/ccbnep

TNRCC

Barry R. McBee, *Chairman*
R. B. Ralph Marquez, *Commissioner*
John M. Baker, *Commissioner*

Dan Pearson, *Executive Director*

Authorization for use or reproduction of any original material contained in this publication, i.e., not obtained from other sources, is freely granted. The Commission would appreciate acknowledgment.

Published and distributed
by the
Texas Natural Resource Conservation Commission
Post Office Box 13087
Austin, Texas 78711-3087

The TNRCC is an equal opportunity/affirmative action employer. The agency does not allow discrimination on the basis of race, color, religion, national origin, sex, disability, age, sexual orientation or veteran status. In compliance with the Americans with Disabilities Act, this document may be requested in alternate formats by contacting the TNRCC at (512) 239-0028, Fax 239-4488 or 1-800-RELAY-TX (TDD), or by writing P.O. Box 13087, Austin, TX 78711- 3087.

printed on recycled paper using soy-based ink

CORPUS CHRISTI BAY NATIONAL ESTUARY PROGRAM

The Corpus Christi Bay National Estuary Program (CCBNEP) is a four-year, community based effort to identify the problems facing the bays and estuaries of the Coastal Bend, and to develop a long-range, Comprehensive Conservation and Management Plan. The Program's fundamental purpose is to protect, restore, or enhance the quality of water, sediments, and living resources found within the 600 square mile estuarine portion of the study area.

The Coastal Bend bay system is one of 28 estuaries that have been designated as an **Estuary of National Significance** under a program established by the United States Congress through the Water Quality Act of 1987. This bay system was so designated in 1992 because of its benefits to Texas and the nation. For example:

- Corpus Christi Bay is the gateway to the nation's sixth largest port, and home to the third largest refinery and petrochemical complex. The Port generates over $1 billion of revenue for related businesses, more than $60 million in state and local taxes, and more than 31,000 jobs for Coastal Bend residents.
- The bays and estuaries are famous for their recreational and commercial fisheries production. A study by Texas Agricultural Experiment Station in 1987 found that these industries, along with other recreational activities, contributed nearly $760 million to the local economy, with a statewide impact of $1.3 billion, that year.
- Of the approximately 100 estuaries around the nation, the Coastal Bend ranks fourth in agricultural acreage. Row crops -- cotton, sorghum, and corn -- and livestock generated $480 million in 1994 with a statewide economic impact of $1.6 billion.
- There are over 2600 documented species of plants and animals in the Coastal Bend, including several species that are classified as endangered or threatened. Over 400 bird species live in or pass through the region every year, making the Coastal Bend one of the premier bird watching spots in the world.

The CCBNEP is gathering new and historical data to understand environmental status and trends in the bay ecosystem, determine sources of pollution, causes of habitat declines and risks to human health, and to identify specific management actions to be implemented over the course of several years. The 'priority issues' under investigation include:

- altered freshwater inflow
- declines in living resources
- loss of wetlands and other habitats
- bay debris
- degradation of water quality
- altered estuarine circulation
- selected public health issues

The **COASTAL BEND BAYS PLAN** that will result from these efforts will be the beginning of a well-coordinated and goal-directed future for this regional resource.

STUDY AREA DESCRIPTION

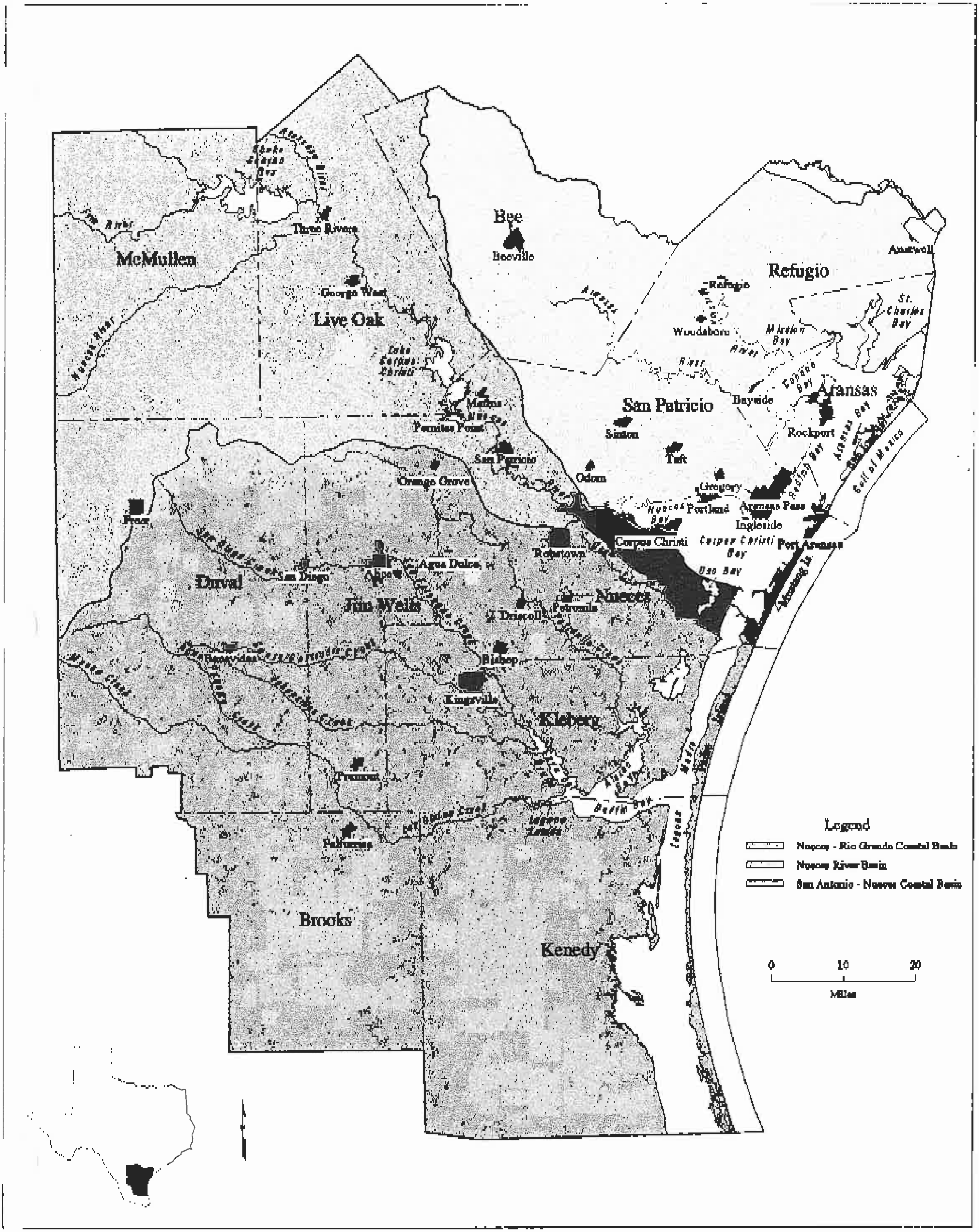
The CCBNEP study area includes three of the seven major estuary systems of the Texas Gulf Coast. These estuaries, the Aransas, Corpus Christi, and Upper Laguna Madre are shallow and biologically productive. Although connected, the estuaries are biogeographically distinct and increase in salinity from north to south. The Laguna Madre is unusual in being only one of three hypersaline lagoon systems in the world. The study area is bounded on its eastern edge by a series of barrier islands, including the world's longest -- Padre Island.

Recognizing that successful management of coastal waters requires an ecosystems approach and careful consideration of all sources of pollutants, the CCBNEP study area includes the 12 counties of the Coastal Bend: Refugio, Aransas, Nueces, San Patricio, Kleberg, Kenedy, Bee, Live Oak, McMullen, Duval, Jim Wells, and Brooks.

This region is part of the Gulf Coast and South Texas Plain, which are characterized by gently sloping plains. Soils are generally clay to sandy loams. There are three major rivers (Aransas, Mission, and Nueces), few natural lakes, and two reservoirs (Lake Corpus Christi and Choke Canyon Reservoir) in the region. The natural vegetation is a mixture of coastal prairie and mesquite chaparral savanna. Land use is largely devoted to rangeland (61%), with cropland and pastureland (27%) and other mixed uses (12%).

The region is semi-arid with a subtropical climate (average annual rainfall varies from 25 to 38 inches, and is highly variable from year to year). Summers are hot and humid, while winters are generally mild with occasional freezes. Hurricanes and tropical storms periodically affect the region.

On the following page is a regional map showing the three bay systems that comprise the CCBNEP study area.



Corpus Christi Bay National Estuary Program Study Area

CORPUS CHRISTI BAY NATIONAL ESTUARY PROGRAM

AMBIENT WATER, SEDIMENT AND TISSUE QUALITY OF CORPUS CHRISTI BAY STUDY AREA:

PRESENT STATUS AND HISTORICAL TRENDS

Project Final Report

Principal Investigators:

George H. Ward
Neal E. Armstrong

Center for Research in Water Resources
The University of Texas at Austin

November 1997

AMBIENT WATER, SEDIMENT AND TISSUE QUALITY OF CORPUS CHRISTI BAY STUDY AREA: PRESENT STATUS AND HISTORICAL TRENDS

Principal Investigators:

George H. Ward
Neal E. Armstrong
Center for Research in Water Resources
The University of Texas at Austin

EXECUTIVE SUMMARY

For many years, data on the physico-chemical quality of water and sediment have been collected in the Corpus Christi Bay system by a variety of organizations and individuals. The purpose of this project was to compile these data, and to perform a quantitative assessment of water and sediment quality of Corpus Christi Bay and its evolution over time. Tissue quality was included as well in the project scope. There were three key objectives:

(1) compilation of a comprehensive data base in machine-manipulable format;
(2) analysis of space and time variation in water, sediment and tissue quality parameters;
(3) identification of probable causal mechanisms to explicate the observed variations.

Their accomplishment provides a foundation for further scientific study of the Corpus Christi Bay system, and for a general understanding of the controls and responses of its water and sediment quality, which must underlie rational management of the resources of the system.

A principal product of this study is the compilation of a digital data base composed of water-quality, sediment-quality and tissue-quality data from 30 data collection programs performed in Corpus Christi Bay. This compilation included data from the three most important ongoing monitoring programs in Corpus Christi Bay: the Texas Natural Resource Conservation Commission (TNRCC) Statewide Monitoring Network, the Texas Parks and Wildlife Department (TPWD) hydrographic observations from its Coastal Fisheries program, and the hydrographic and biochemical data of the Texas Department of Health Seafood Safety Division program. The important surveys and research projects sponsored by the Texas Water Development Board (TWDB) and maintained in its digitized Coastal Data System are included. Several recent federal data-collection projects are represented, namely those of the U.S. Corps of Engineers (USCE) Galveston District, Environmental Protection Agency (EPA), National Ocean Service, and U.S. Fish & Wildlife Service.

This compilation also entailed keyboarding of other major data sets, many of which exist in limited hardcopy and are virtually unobtainable, including the U.S. Corps of Engineers Galveston District water and sediment surveys of the 1970's, data of the Texas Game Fish & Oyster Commission from the 1960's, the Reynolds-sponsored "baseline" surveys of the early 1950's, the Submerged Lands Project of the Bureau of Economic Geology, and the data collections by the now-defunct Ocean Science and Engineering Laboratory of Southwest Research Institute. Other entries in this compilation include research projects whose data are published only in limited technical reports or academic theses, all of which were keyboarded. A major data compilation effort of the project was devoted to determination of latitude/longitude coordinates based upon historical sampling station location information, so that all of the data could be unambiguously georeferenced. In addition to supporting the spatial-distribution analyses of this study, this georeferencing data will facilitate incorporation of the data base into geographical information systems.

All told, the digital compilation is the most extensive and detailed long-term record of water and sediment quality ever assembled for Corpus Christi Bay. The study area for this compilation and analysis extends from the landbridge of the Laguna Madre to the southern limit of San Antonio Bay, and includes Baffin Bay, Corpus Christi Bay proper, the Aransas-Copano system, and Mesquite/Ayres Bay. We refer to Aransas, Copano and their secondary systems (including Mesquite) as the upper bays, and to Baffin Bay and the Upper Laguna Madre as the lower bays. The entire CCBNEP study area is referred to as the Corpus Christi Bay "system," to differentiate it from Corpus Christi Bay proper, unless it is clear in context that the CCBNEP study area is intended (such as the first sentence of this paragraph). The complete data base approaches half a million independent records of which water:sediment:tissue are in the approximate ratios 100:10:2, and about 43% of the water-phase data are the "field" parameters temperature/salinity/pH/dissolved oxygen. Each measurement record includes the date, sample depth, latitude and longitude of the sample station, measured variable, estimated uncertainty of measurement expressed as a standard deviation, and a project code identifying the origin of the data. (For tissue data, the sample depth field is replaced by a code identifying the organism.)

The extant period of record for Corpus Christi Bay, with adequate continuity for trends analysis, extends back only to about 1965, except for some traditional parameters and for certain areas of the bay, for which the record can be extended back to the 1950's. As salinity and temperature are the most easily measured variables, they represent the densest and longest data record. For metals and for complex organics, the period of record may extend back only a decade or so. Many of these measurements are below detection limits. For sediment, the data base is even more limited, amounting to one sample per 50 square miles per year, and extending back in time at most to the 1970's.

Spatial aggregation of the data was accomplished by two separate segmentation systems for Corpus Christi Bay, the TNRCC Water Quality Segmentation of 27 segments, and a system of 178 hydrographic segments devised by this project and designed to depict the effects of morphology and hydrography on water properties.

(The 27 TNRCC segments include the original 15 specified by the Scope of Work, to which we added 5 classified segments and 7 unclassified.) For each segment of both systems, detailed statistical analyses were performed of 109 water-quality parameters and 83 sediment-quality parameters, in addition to supplementary, screened, or transformed variables. Each statistical analysis included basic sampling density information, means and standard deviations, with three different treatments of measurements below detection limits (BDL), and a linear trend analysis over the period of usable record, with confidence limits on the slope. Therefore, statistical analyses addressing water/sediment quality were performed of about 200 parameters in about 200 different segments, a total of about 40,000 independent statistical analyses, since each parameter/segment comprises an independent data set. For tissue data, an even more extensive suite of analytes were compiled, but the statistical analyses were confined to a subset of these analytes because of the sparsity of the data base. In addition to parameter differentiation, tissue data had to be further separated according to organism, portion of organism analyzed (i.e., whole versus fileted), and reporting by dry- or wet-weight, each combination of which represented an independent statistical analysis.

A summary of the findings on the water and sediment quality "climate" begins with salinity, which acts as a water mass tracer and general habitat indicator for Corpus Christi Bay. In contrast to the estuaries on the upper Texas coast, salinity gradients across Corpus Christi Bay from the sea to the regions of inflow are on average rather flat. The most substantial gradient of salinity is, rather, from north to south, from Copano Bay to Baffin Bay, a combined result of diminishing inflow with distance to the south and increasing evaporation. Mean salinities often exceed seawater concentrations, sometimes by large amounts, especially in the lower bays (the Upper Laguna and Baffin Bay). Vertical salinity stratification of bay waters is slight by estuarine standards, generally averaging less than 0.6 ppt/m. There is no apparent correlation between mean salinities and ship channels, suggesting that density currents as a mechanism of salinity intrusion are rarely important in Corpus Christi Bay. While freshwater inflow is the ultimate control on salinity, inflow proves to be a poor statistical predictor of individual measurements of salinity, even with long-term averaging of the antecedent inflow. This illustrates that the variability of salinity is influenced by factors other than simply the level of inflow.

In the bays more influenced by freshwater inflow, *viz*. Copano Bay, the main body of Corpus Christi Bay and Nueces Bay, there has been a general increase in salinity over the three-decade period of record, on the order of 0.1 ppt per year. During the same period there has been a declining trend in monthly-mean inflow to these same bays, over 50% in Corpus Christi and Nueces Bays, less in Copano (which also logged a smaller increase in salinity). Our favored hypothesis is that this decline in mean inflow, which appears to be due to diminishing frequency and magnitude of freshets, is responsible for the increase in salinity. No clear trends in salinity emerged for the Upper Laguna or Baffin Bay.

The principal variation in water temperature in Corpus Christi Bay is the seasonal signal, ranging about 14 to 30°C from winter to summer, which means that temperature is primarily controlled by surface fluxes, especially the seasonal heat budget, and much less by peripheral boundary fluxes and internal transports. The horizontal gradient across the study area is from north to south, ranging 2-4°C. There is little systematic stratification, on average a slight upward increase on the order of 0.1°C/m, due to near-surface heat absorption. Over the three-decade period of record, water temperature in the upper bays and main body of Corpus Christi Bay, especially in the open waters, has declined at a nominal rate of 0.05°C/yr. There are no clear trends in the lower bays. It is interesting to note that the same decline in temperature, at approximately the same rate, was discovered in Galveston Bay (Ward and Armstrong, 1992a).

TNRCC applies a 35°C temperature standard uniformly to the entire Texas coast, without cognizance of the natural gradient of increasing temperatures toward the south (a gradient to which the indigenous organisms would have presumably acclimated). The shallow, poorly circulated sections of the Corpus Christi system are most prone to higher temperatures, especially those in the lower bays, and exceedances of the TNRCC 35°C standard occur at a low rate–a couple of percent–mainly in summer. Only two regions have substantially higher frequencies of exceedance, in Nueces Bay and Oso Bay, both affected by return flows from power plants. Given this low frequency of exceedance coupled with the general decline in water temperatures over time, we conclude that hyperthermality is not a problem in Corpus Christi Bay.

Dissolved oxygen (DO) consistently averages near (and above) saturation throughout the CCBNEP study area, with frequent occurrence in the data record of substantial supersaturation. Exceptions to this are in poorly flushed tributaries and areas influenced by wasteloads, especially the Inner Harbor. The predominant variation in DO is due to seasonal changes of solubility. In the open, well-aerated areas of the bay, vertical stratification is slight, averaging on the order of 0.1 ppm/m, and is considered to be the result of DO aeration at the surface in concert with water-column and sediment biochemical oxygen consumption.

We examined episodic occurrence of low DO's. Hypoxia (which we define to be DO ≤ 2 ppm) is rare, occurring at most in several percent of the data in a minority of regions of the bay, and primarily in measurements near the bottom in deeper water. The exception is the Inner Harbor, where hypoxia has occurred more frequently, in about one-fourth of the measurements, but still primarily near-bottom. Near-zero DO (defined to be DO ≤ 0.5 ppm) is rarer yet, representing perhaps half of the hypoxia events, mainly confined to the Inner Harbor and Nueces River.

Most areas of the bay have a relative frequency of DO below the TNRCC standard of 5 ppm (without vertical averaging, diurnal-excursion allowance, or screening by flow, as required for its direct applicability) of a couple of percent, almost always in the summer or early fall. There are scattered higher frequencies of violations, especially in proximity to sources of inflow and wasteloads, and in the shallow, poorly-circulating areas near the barrier island, especially in the Upper Laguna. The apparent contradiction between the observation that the system is at or above saturation much of the time, and yet has a nonnegligible frequency of standard violation, 10-20% in a few areas, is reconciled by noting that much of the year DO solubility falls very close to the standard, because of the high natural tempera-tures and salinities in this area. This also implies that the clearance between solubility and a level of DO stress is small, so under these conditions the assimilative capacity for oxygen demands may be limited.

Conventional water-phase organic contaminants as measured by biochemical oxygen demand (BOD), oil & grease, volatile suspended solids (VSS) and volatile solids, are generally highest in the Inner Harbor. The frequency of measurement of these parameters has declined substantially in recent years, and trends are therefore uncertain. In the open waters of Corpus Christi Bay, BOD seems to be declining, and wherever adequate data for analysis exist, VSS is declining, probably the result of improved waste treatment.

Like all of the Texas bays, Corpus Christi is turbid. Long-term average total suspended solids (TSS) range 20-100 ppm throughout most of the study area, higher in Nueces, Copano and Corpus Christi Bay, as well as in Baffin. Stratification in TSS is noisy, but on the order of 5 ppm/m declining upward, which is consistent with settling of larger particles to the bottom as well as a near-bottom source of particulates from scour of the bed sediments. The highest TSS concentrations and highest stratification are found in Nueces Bay.

The most remarkable feature of TSS is its decline over time, increasing in significance from north to south across the study area, at a rate sufficient to have reduced the average concentration by about 25% in the upper bays and by about 50% in the lower bays over the last two decades. This could be caused by several factors, including a general reduction of TSS loading to the bay or altered mobilization within the bay system itself. Suspended sediment is an intrinsic and important aspect of the Corpus Christi Bay environment; its decline is not necessarily beneficial.

Nitrogen and phosphorus nutrients in the water column are highly variable. Ammonia nitrogen is generally higher in regions affected by waste discharges, especially the Inner Harbor, while nitrate nitrogen and phosphorus are typically highest in regions affected by runoff and inflow. Concentrations of inorganic nitrogen are about 0.1 ppm, except much higher, around 0.5 ppm, in Copano and in the Inner Harbor (the latter due to high ammonia). Total phosphorus is about 0.05 ppm, except around 0.15 ppm in Nueces Bay, Copano Bay and Baffin Bay. These mean concentrations are more-or-less typical of other Texas bays (e.g., Longley, 1994), though total inorganic nitrogen is about half the levels found in Galveston Bay and total phosphorus is about one-fourth (Ward and Armstrong, 1992a). Generally where the nitrogen species are high in concentration, they exhibit a declining trend. No clear trends are apparent in the phosphorus data. In the sediment phase, highest concentrations of Kjeldahl nitrogen occur in the Inner Harbor region, and the Upper Laguna. Sediment phosphorus is relatively uniform throughout the system.

Generally water-phase total organic carbon (TOC) values decrease southward from 20-30 ppm in Copano to 5-15 ppm in Baffin and the Laguna, with a seasonal peak in early summer. Larger values (by an order of magnitude) occur in the Inner Harbor. Sediment TOC distributions generally run counter to the water phase, increasing southward across the study area with the lowest values of sediment TOC in the Inner Harbor. Nueces Bay shows substantially depressed values of TOC in both water and sediment. There is a widespread declining trend in water-phase TOC at a rate that would reduce concentrations by about one-fourth over two decades. (The prominent exception to this is in the Inner Harbor, where average water-phase TOC is the highest in the study area, and is increasing in time.) Where sufficient sediment TOC data exist to establish a trend, this trend generally is also declining in time. Unfortunately, the data for chlorophyll-a is too sparse and noisy to determine whether any correlated time trends occur in it as well, so we cannot judge whether the decline in TOC is due to reduced primary production or to reduced loadings.

Contaminants such as coliforms, metals and trace organics show elevated levels in regions of runoff and waste discharge, with generally the highest values in the Inner Harbor, and low values in the open bay waters. The highest average coliforms in the system occur in the nearshore segments of Corpus Christi Bay from Corpus Christi Beach to Oso Bay. Nueces Bay is a region consistently high in metals, in both the water column and the sediment, as are Baffin Bay, Copano Bay, a region of the Upper Laguna around the Bird Islands, the La Quinta Channel, and Redfish Bay near Aransas Pass. The metals copper, nickel and zinc, in particular, have elevated concentrations in water generally throughout Corpus Christi Bay. The water-phase metals data were so sparse and noisy in time that reliable trends could not generally be established. With respect to sediment metals, arsenic, cadmium, mercury, and zinc are elevated, with concentrations generally on the same order as Galveston Bay. Inner Harbor sediment metals are similar to the upper Houston Ship Channel except zinc, for which its sediments are an order of magnitude *higher* than those of the Houston Ship Channel. This raises the speculation of whether the Inner Harbor could be the ultimate source for elevated zinc in the system. For sediment metals in the principal components of the system, where a trend can be reliably established it is generally declining. An exception to the general declining trends is sediment zinc, for which widespread possible increasing trends are indicated in large areas of the open waters of Corpus Christi Bay and Baffin Bay.

No definitive statements can be made about water-phase semi-volatile organics such as pesticides and PAH's, because data is sparse, and very few measurements are uncensored, most being simply reported as below detection limits. For example, the best-monitored pesticide is DDT, for which most areas of the bay do not have data. Only four non-zero average values occur in the entire study area, two in Ayres Bay, and one each in Nueces Bay and Baffin Bay. For toxaphene, only one non-zero value occurs, in Nueces Bay. The situation is similar for the other organics, with only one or a few non-zero values, and inadequate data to determine any trends or spatial variation.

The situation is a little better for sediment-phase data, but still most of the system is unsampled, and much of the data which does exist is below detection limits. The highest concentrations of the common pesticides are found in Baffin Bay and

Copano Bay. Concentrations of sediment pesticides in Nueces Bay are not especially high, except for toxaphene. PCB's and PAH's exhibit very high concentrations in the Inner Harbor. Elevated concentrations of PCB's also occur in Redfish Bay. There are consistent elevated concentrations of some of the PAH compounds in Nueces Bay, Copano Bay, and Mesquite Bay, but not in the Upper Laguna.

Considering the effort required to obtain, digitize and compile the tissue data for the CCBNEP study area, the information yield is disappointing. Pooling and analysis of the data are hampered by the noncomparable attributes of organism sampled, portion of organism analyzed (whole versus edible portions), and reporting convention (wet-weight versus dry-weight), in addition to the usual discriminants of analyte and geographical position. The most-sampled organism is the American oyster, with most samples from Nueces and Aransas Bays, followed by the blue crab, speckled trout, red drum and black drum. There is one sample each of brown shrimp and white shrimp. By far, the greatest quantity of analyses have been performed for the metals. For the oyster, Nueces Bay and Copano Bay exhibit systematically elevated metals in the tissue, Nueces Bay having the highest mean tissue concentrations in the study area for cadmium, copper, lead and zinc. This conclusion generally agrees with the relative concentrations in the sediments, if the Inner Harbor and tertiary bays are discounted. Blue crab data in Redfish Bay and Baffin Bay show elevated levels of most metals. Statistical analysis of the black drum data base was possible only for Nueces Bay, which indicated some elevated metals concentrations, especially for mercury and zinc, and when a time trend could be resolved, it is increasing. The data base of *detected* PAH's and related hydrocarbons is negligible. For only a few, such as pyrene, have there been detects logged in the data.

From a systemic point of view, the most significant potential problems affecting the study area as a whole are suspended particulates, nutrients and salinity. As summarized above, declines of TSS, inorganic nitrogen and TOC were found. More data is needed to determine whether there is a decline in productivity, or what the optimum levels are for Corpus Christi Bay. Salinity of Corpus Christi Bay has been a major source of controversy, especially within the past decade, because of its perceived value as a habitat indicator that also measures freshwater inflow. At this writing, the City of Corpus Christi water supply in the Nueces River reservoirs is threatened by a drought, and the conflict between human water-supply requirements and the needs of the estuary ecosystem has been brought into sharp relief. One result of the present study, disclosure of increasing salinity that seems to be associated with declines in mean inflow, suggests that salinity will continue to be at the center of management issues and strategies for this system, even after the current drought has abated.

Several deficiencies of this data set are noted, as they relate to the interpretation of water and sediment quality. Adequacy of a data base is judged relative to the ability to resolve the various scales of variation, and in this respect Corpus Christi Bay is undersampled. An estuary such as Corpus Christi Bay is subject to a variety of external controls, all of which contribute to variation in space and time. The intermixing of fresh and oceanic waters imposes spatial gradients in both the horizontal and the vertical. The effects of tides, meteorologically driven circulations, and transient inflows all contribute to extreme variability in time. Superposed upon all of this are the time- and space-varying influences of human activities.

Despite the hundreds of thousands of separate measurements compiled in this study, from extensive and overlapping routine monitoring and survey programs by several state agencies and numerous special surveys, when these data are subdivided by specific parameters, each of which measures a different aspect of the water/sediment quality "climate," aggregated by region of the bay (segments) and distributed over time, the data record is seen to be rather sparse. Continuity in space is undermined by too few stations, and by inconsistency in the suite of measurements at different stations. Ability to resolve long-term trends in the face of high intrinsic variability requires data over an extended period. Continuity in time is undermined by infrequent sampling, and the replacement of one parameter by another without sufficient paired measurements to establish a relation.

After a relative peak in the mid-1970's, data collection, as reflected in the number of sampling programs underway and the density of the network of stations, has been declining. Considering that Corpus Christi Bay is undersampled, this trend is in the wrong direction. (There are some exceptions: there have been recent increases in salinity and DO sampling, mainly due to the activities of TPWD, and in trace constituent sampling, due to increased concern with metals and organic toxicants and to the advancement of instrumental analysis.) To maintain a monitoring project within limited resources requires a compromise between station density, temporal frequency, and the extent of the suite of analytes. Cost for all three have been increasing, the last due to more precise and expensive laboratory methodologies. There is no doubt that economics is one of the prime factors forcing the recent decline in all of these, especially in the spatial and temporal intensity of sampling. That does not mitigate the fact that our ability to understand and manage Corpus Christi Bay is concomitantly diminished.

Data management is generally poor. Both modern and historical data bases have been compromised in various ways. Too many entries in the data record had to be excluded from the analyses presented here because the data were unreliable. It is our belief that much of this unreliability was not introduced in the original measurement but in the subsequent handling of the data. The most pressing management problem for historical data in the Corpus Christi area, as well as in other areas of the Texas coast, is preservation of the older data. Much irreplaceable and invaluable information on the Corpus Christi Bay system has been lost.

Recommendations are offered for data-collection procedures, data management, historical data management, and specific research topics.

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

TABLE OF CONTENTS
(continued)

LIST OF TABLES

* Parameter abbreviations given in Table 2-1

LIST OF TABLES
(continued)

LIST OF TABLES
(continued)

LIST OF TABLES
(continued)

LIST OF TABLES
(continued)

LIST OF FIGURES

* Parameter abbreviations given in Table 2-1

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

ABBREVIATIONS AND ACRONYMS
(See also Table 2-1 and Section 8.1)

AA	atomic absorption
AE	atomic emission
APHA	American Public Health Association
BaP	benzo(a)pyrene
BDL	below detection limits
BEG	Bureau of Economic Geology, University of Texas
BOD	biochemical oxygen demand
BOGAS	Based On Graphical or Arithmetical Suppositions
CAFO	confined animal feeding operation
CBI	Conrad Blucher Institute of TAMU–Corpus Christi
CCBNEP	Corpus Christi Bay National Estuary Program
CC-NCHD	Corpus Christi-Nueces County Health Department
CCSC	Corpus Christi Ship Channel
CD	compact disc
CD-ROM	compact disc-read-only medium [memory]
CDS	Coastal Data System of TWDB
cfs	cubic feet per second
COD	chemical oxygen demand
CP&L	Central Power and Light
DDT	dichlorodiphenyltrichloroethane
DL	detection limit
DO	dissolved oxygen
EM	electromagnetic
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency, also USEPA
EqP	Equilibrium Partitioning
FTU	Formazin Turbidity Unit
GBNEP	Galveston Bay National Estuary Program
GC-MS	gas chromatograph-mass spectrometer
GIWW	Gulf Intracoastal Waterway
GOM	Gulf of Mexico
GPS	Geographical [Geodetic, Global] Positioning Systems [Satellites]
HDR	Henningson, Durham and Richardson, Inc.
HSPF	Hydrological Simulation Program FORTRAN
TU	Jackson Turbidity Unit
LORAN	Long-Range [Radio] Navigation
LS	least squares (referring to statistical regression)
MGD	million gallons per day
MW	megawatts
NGVD	National Geodetic Vertical Datum
NIST	National Institute of Standards and Technology
NOAA	National Oceanic and Atmospheric Administration
NOS	National Ocean Service of NOAA
NOS S&T	National Ocean Service Status and Trends Program

ABBREVIATIONS AND ACRONYMS
(continued)

NTU	Nephelometric Turbidity Unit
OLS	ordinary least squares (referring to statistical regression)
O&M	Operations and Maintenance
PAH	polycyclic aromatic hydrocarbon
PCB	polychlorinated biphenyl
PI	Principal Investigator
POR	period of record
QA/QC	quality assurance and quality control
SEDxxxxx	general abbreviation for sediment quality analyte, see Table 2-1
SEE	standard error of the estimate
SES	steam-electric station
SMN	Statewide [Stream] Monitoring Network
SWRI	Southwest Research Institute
TAMU	Texas A&M University
TDH	Texas Department of Health
TDWR	Texas Department of Water Resources
TGFOC	Texas Game, Fish and Oyster Commission
TXxxxxx	general abbreviation for tissue quality analyte, see Section 8.1
TNRCC	Texas Natural Resource Conservation Commission
TOC	total organic carbon
TPWD	Texas Parks & Wildlife Department
TSS	total suspended solids
TWC	Texas Water Commission
TWDB	Texas Water Development Board
USCE	U.S. Corps of Engineers, U.S. Army Corps of Engineers, U.S. Army Engineer
USEPA	U.S. Environmental Protection Agency, also EPA
USF&WS	U.S. Fish and Wildlife Service
USGS	U.S. Geological Survey
VSS	volatile suspended solids
WQxxxxx	general abbreviation for water quality analyte, see Table 2-1

1. INTRODUCTION

1.1. Background and project motivation

The Corpus Christi region could be characterized as geographically and culturally transitional. It is near the northern limit of the chaparral brush country and the southern limit of the coastal-plain grassland prairies. Corpus Christi has been a prime terminal supporting the vast coastal-plain ranches of south Texas, but is far enough north to be influenced by the planters from the east. It is also far enough south to be affected by Spanish traditions. The commerce of the port in the late Nineteenth Century included exports of beef and cotton, and imports of fruits and metals from Mexico. Much of this cultural overlap is tied to the transitional hydroclimatology of the region. It is tropical much of the time, but still far enough north to be influenced by the midlatitude westerlies and their synoptic disturbances which fling cold air masses down the Great Plains. It is also usually arid, a result of the eastward extension of the rainshadow of the North American cordillera, and the infrequency of midlatitude disturbances intruding this far south. But the exceptions to this aridity are extreme: freshets on the Nueces and diluvial tropical storms, either of which can result in flooding and render much of the bay fresh. This combination of arid, tropical conditions with intermittent intense rainfall has allowed vigorous agricultural production in the area. These extremes in hydroclimatology are also the primary external forcings of Corpus Christi Bay that ultimately govern its quality.

For Corpus Christi Bay and the adjacent systems of Aransas-Copano Bay and Laguna Madre, concerns about the quality of the system have arisen rather more recently than the urbanized and industrialized bays on the upper Texas coast. In 1841, when the City of Houston was passing an "anti-pollution" ordinance prohibiting the accumulation of sawdust on the shoreline of Buffalo Bayou (Sibley, 1968), the dominant concern of the Kinney-Aubrey trading post on Corpus Christi Bay was defending against raids by Indians and Mexicans. The boom of prosperity in Texas in the last quarter of the Nineteenth Century expressed itself in the Corpus Christi area as expansions in ranching, agriculture, and commercial fishing, in synergism with incursions of railroads and shipping, the trade stimulated by all of these activities, and, of course, tourism. But the population increase attending this expansion was modest in comparison to that of the upper coast. By 1900, Houston and its port had become a major industrial center, while Corpus Christi was regarded as primarily a tourist resort.

Urbanization and industry are relative latecomers to the area. In the early years of this century, while the effects of industrial pollution were being logged in the upper bays (TGFOC, 1928, described the impacts of petroleum operations on Galveston Bay in 1920: "Fishing in the ship channel was ruined, and most of the marine life had been driven from the upper portions of the bay. Bathers often received generous coatings of oil."), an example of industrial concern in Corpus Christi was the loss of the turtle fishery and the canneries it supported. Only in the 1930's did heavy industry begin to be situated on the industrial canal, with construction of the Southern Alkali Corporation plant (which used oyster shell from Nueces Bay). Oil production began in this same decade of the 1930's near White Point and in the Saxet Field, which stimulated shipping and later refining, and was the major impetus for growth in the area.

For Corpus Christi Bay and the adjacent systems of Aransas-Copano Bay and Laguna Madre, concerns about the quality of the system have arisen rather more recently than for the urbanized and industrialized bays on the upper Texas coast. Up to World War II, there appear no reports or indications of perceived pollution problems, in contrast to the upper coast. (Outbreaks of shellfish poisoning occurred in Galveston Bay in 1944, due to sewage contamination, Wise et al., 1944, 1948. TGFOC, 1946, states "The total catch from the Galveston area is an insignificant per cent of the total production in Texas waters and can be expected to remain so until the heavy industrial pollution of that region is abated.") Far more fish kills have occurred in the Corpus Christi Bay system due to freshets and freezes than to contamination. In the last two decades public attention and concern for the Corpus Christi Bay system has changed. With accelerating urban development, awareness of the potential impacts on the bay has increased, and maintenance of the health of the system–and its reconciliation with goals of municipal growth and industrial development–has become a major issue. With this concern is the recognition that the quality of Corpus Christi Bay must be managed.

The cornerstone of management of a natural system like Corpus Christi Bay is the ability to determine responses of the system to changes in external or controlling factors, i.e. its "controls," in the form of cause-and-effect relations. Qualitative, anecdotal information is interesting from a historical viewpoint, but does not contribute to answering the questions of whether significant problems in water quality presently exist in Corpus Christi Bay and whether there is (or has been) a long-term alteration in water quality. Two elements are needed in order to appraise variation in water quality and to identify its cause. First is a quantitative measure, i.e. identification and analysis of a parameter (or parameters) indicative of water quality, which in principle can provide time-space continuity. Complaints of declines in a fishery, for example, are dramatic evidence of something, but offer little basis for scientific evaluation. Instead, a physical or chemical parameter (or several, or many) is needed upon which the viability of that fishery depends, and which represents the impacts of some natural or human process on waters of the bay. The second element needed is an extensive data base on the parameter. This, of course, is the real obstacle.

The data base must have sufficient spatial and temporal resolution to sample the variability of the parameter. The data base must also extend over a considerable period of time, for two reasons. First, from the standpoint of determining time trends, a long period of record is absolutely indispensable. This is intuitive, but can be demonstrated mathematically. From a statistical point of view, one will recall that the variance of the sampling distribution of the slope of a linear least-squares trend line is:

$$\sigma_r^2 / (N \sigma_t^2)$$

in which t is the independent variable (time) with variance σ_t^2, N is the number of data points in the estimate and σ_r^2 is the variance of the dependent variable (concentration of some indicator parameter, in our case) about the trend line. (For a uniform sampling rate, this becomes exactly $12\sigma_r^2/N^4$.) The variance of t is directly related to the length of the period of measurement, as well as to the uniformity of distribution of measurements in that period. This relation tells us that the statistical confidence in the trend and the explained variance of the trend will each be proportional both to the number of data points N and the variance of the independent variable σ_t^2. Too short a period of record, or too few measurements in time, or especially both, will hamper one's ability to establish a trend with any statistical confidence. If there is no trend line, but one is seeking a harmonic signal in the data, the variance of the estimated amplitude (and, hence, both confidence in its determination and the statistical predictive power of the model) is proportional to $1/(S\ N^3)$, where S is the signal-to-noise ratio (see Walker, 1971). This means that a long series (N) is much more important than a strong signal (S). Or, putting it another way, a sufficiently long series will allow one to discern even weak signals.

Second, a long period of data collection is necessary to encompass the range of variation of external conditions to which the bay is subject and which contribute to the variance in the parameters. Periods of drought and floods, the occurrence of freshets and storms, high-water stands and low-water events, great- and small-declination tides, winds and waves, all contribute to the range of variation in an observed parameter. The mathematical demonstration is the same as that above, except the independent variable is not time but any proposed forcing function such as river flow or waste-loading rate. Only a long period of record assures that the full range of conditions is reflected in the data base. Too short a period will limit the range of variation of the forcing variable, as well as run the risk of biasing the data.

Generally, any single data-collection program lacks the resources and longevity to develop a data base sufficiently comprehensive for analysis of water quality levels and trends in a system such as Corpus Christi Bay. This is due to the extreme natural variability of the water-quality parameters. The best prospect for a definitive study is to begin with a synthesis of data from a number of programs, using the entire spatial and temporal scope of each program. This is the strategy followed here.

1.2 Objectives and prosecution of project

For many years, data relating to the quality of water and sediment have been collected in the Corpus Christi Bay system by a variety of organizations and individuals. The objectives of data collection have been equally varied, including the movement and properties of water, the biology of the bay, waste discharges and their impacts, navigation, geology and coastal processes, and fisheries. While the specific purposes of the individual data collection projects have limited each project in time and space, the data have great potential value to the Corpus

Christi Bay National Estuary Program (CCBNEP) if they can be combined into a comprehensive data base yielding a historical depiction of the quality of the bay environment.

The purpose of this project was to compile and evaluate these data, and to employ these data in a quantitative assessment of water and sediment quality of Corpus Christi Bay and its evolution over time. There were several subordinate objectives in the project, as outlined in the following sections. However, the key objectives were threefold, *viz.*:

(1) compilation of a comprehensive data base in machine-manipulable format,

(2) analysis of time and space variation (including "trends") in quality parameters,

(3) identification of possible causal mechanisms to explicate the observed variations.

Securing these objectives will provide a foundation for further scientific study of Corpus Christi Bay, for identifying and prioritizing specific problems affecting the quality of the Bay, for formulation and specification of future monitoring programs for the Bay, and for a general understanding of the controls and responses of Bay water quality, which must underlie rational management of the resources of the system.

This project was prosecuted according to the Work Plan (Ward and Armstrong, 1994), which in turn generally conformed to the outline of the draft Scope of Work prepared by CCBNEP management. Two separate documents have been developed from this project, because there are two different aspects of the project that would benefit by independent reporting. The present report employs the data base to characterize the Corpus Christi Bay system, including statistical analyses of the data for key water-quality areas and all TNRCC segments, identification of water-quality problems, and an analysis of apparent mechanisms for the variation in space and time of water quality, and for the occurrence of water-quality problems. This report presents the rationale and formulation of the aggregated data base, including the specification of the water-quality areas. This report also provides an assessment of the historical data base and the data collection programs that have produced it, with gaps and inadequacies identified, and specific recommendations for future monitoring programs in Corpus Christi Bay. A companion report (Ward and Armstrong, 1997a) addresses the data base itself, documenting the sources for the data, formatting of the data, methodology, data quality, and spatio-temporal coverage. This report should function as a Users Guide to the data base, to form the foundation for use of the data base by other researchers. In addition, an abridged version of the present report has been published as the project Summary Report (Ward and Armstrong, 1997b), a format which should satisfy the needs of most readers.

The focus of this study was on the quality of water and sediments in the Corpus Christi Bay system. "Quality" is a broad term, referring in general to any quantitative parameter (or suite of parameters, taken collectively) that can serve as an *indicator* for a potential use of water in Corpus Christi Bay. "Use" in this context means "function" and includes the uses of nature as well as the activities of man. By this definition, "quality" would range from physical properties such as current velocity to organisms of the bay, and would include the atmospheric and terrestrial environs. We adopted a narrower definition, consistent with the draft Scope of Work and with the other projects within the CCBNEP that are examining other components of Corpus Christi Bay, that "quality" is defined by physical, chemical and microbiological constituents associated with the Bay waters or its suspensions.

The study area for this project encompasses the estuarine and coastal nearshore areas extending from the mud flats (a.k.a. middle ground, a.k.a. landbridge, a.k.a. landcut) of the Laguna Madre to the southern limit of San Antonio Bay, and includes Baffin Bay, Corpus Christi Bay proper and its secondary embayments, the Aransas-Copano system, and Mesquite/Ayres Bay, see Figs 1-1 through 1-4. As noted in Chapter 3, we have further enlarged the area of study to include tidal and above-tidal reaches of the principal tributaries, the southern segment of San Antonio Bay including Hynes Bay, and the adjacent Gulf of Mexico out to about the 10-fathom contour. This large study area creates a terminological problem. We refer to the entire study area as the Corpus Christi Bay system. The northern bays of Aransas and Copano, and their respective secondary systems, St. Charles, Mission, Ayres, etc., are referred to as the upper bays. Similarly, the southern elements of the system, Baffin Bay and the Upper Laguna Madre are referred to as the lower bays. An important subregion is Corpus Christi Bay *per se*, which can be confused with the overall study area. We attempt to minimize this confusion by referring to the CCBNEP study area as the "system", and Corpus Christi Bay proper as the "main body", or some similar qualification. Hopefully, the context will prevent confusion.

This project sought data from various sources, relating to the general categories of indicator variables (bacteriological and chemical), nutrients, heavy metals, pesticides, organics, suspended matter, and hydrographic variables, including density and salinity, and created a computer-manipulable data base. Generally, the first portion of the project effort concentrated upon acquisition and transmittal of data holdings, and the latter portion with data entry, and the development of the data base. An important task of this portion of the project was that of recovering historical data sets from oblivion, and one major product of this project is consistent, digital forms of the major water/sediment programs from the Bay.

One of the difficulties encountered in the earlier Water Quality Status and Trends Project carried out as part of the Galveston Bay NEP (Ward and Armstrong, 1992a) was the poor response from many of the major agencies responsible for historical data collection in Galveston Bay. As might be anticipated, we encountered the same problem of poor response again. Time periods of months were required to obtain digital copies of data bases; in a few instances over a year elapsed between our (first) request and our eventual receipt of data. In the

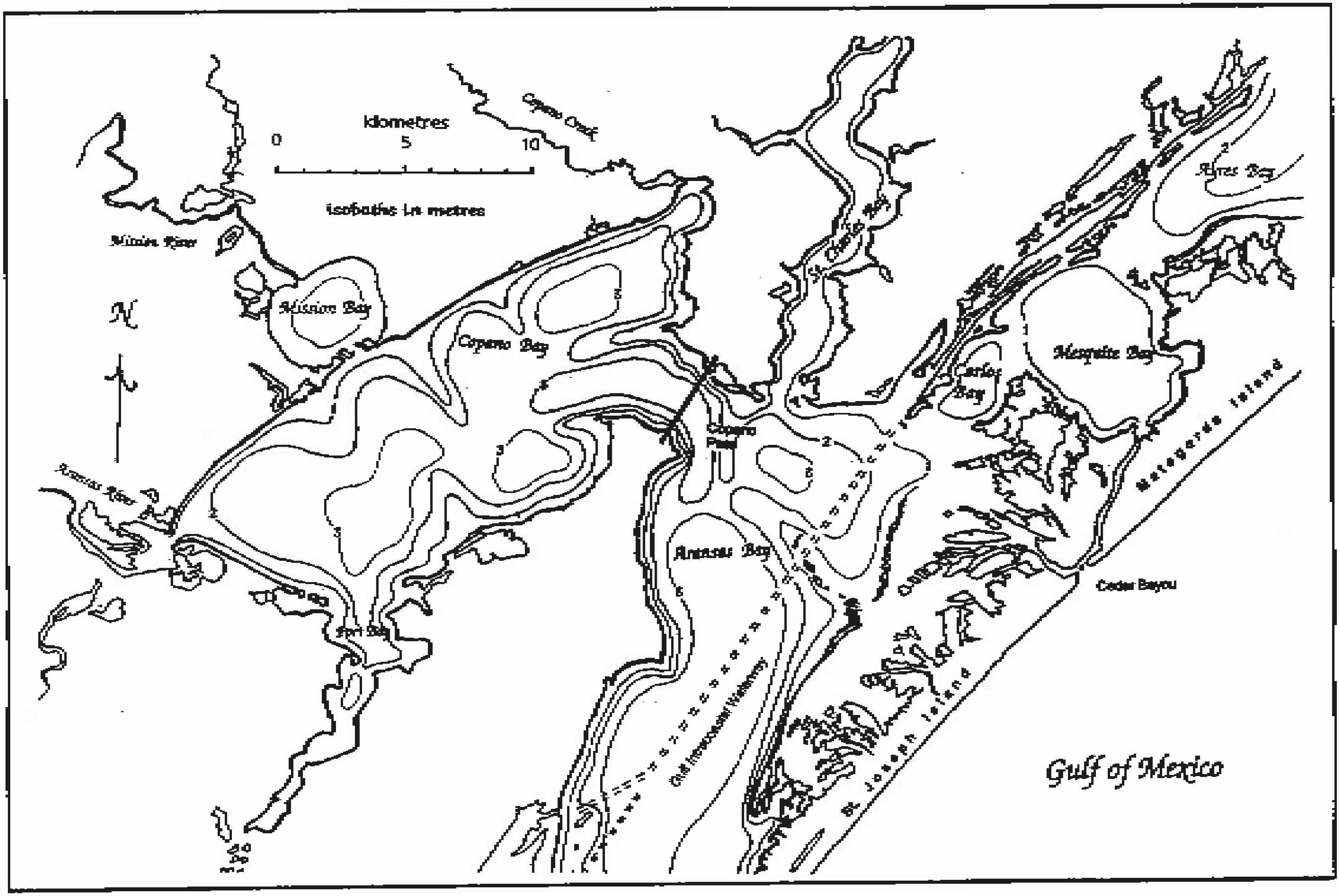


Figure 1-1. Aransas-Copano Bay

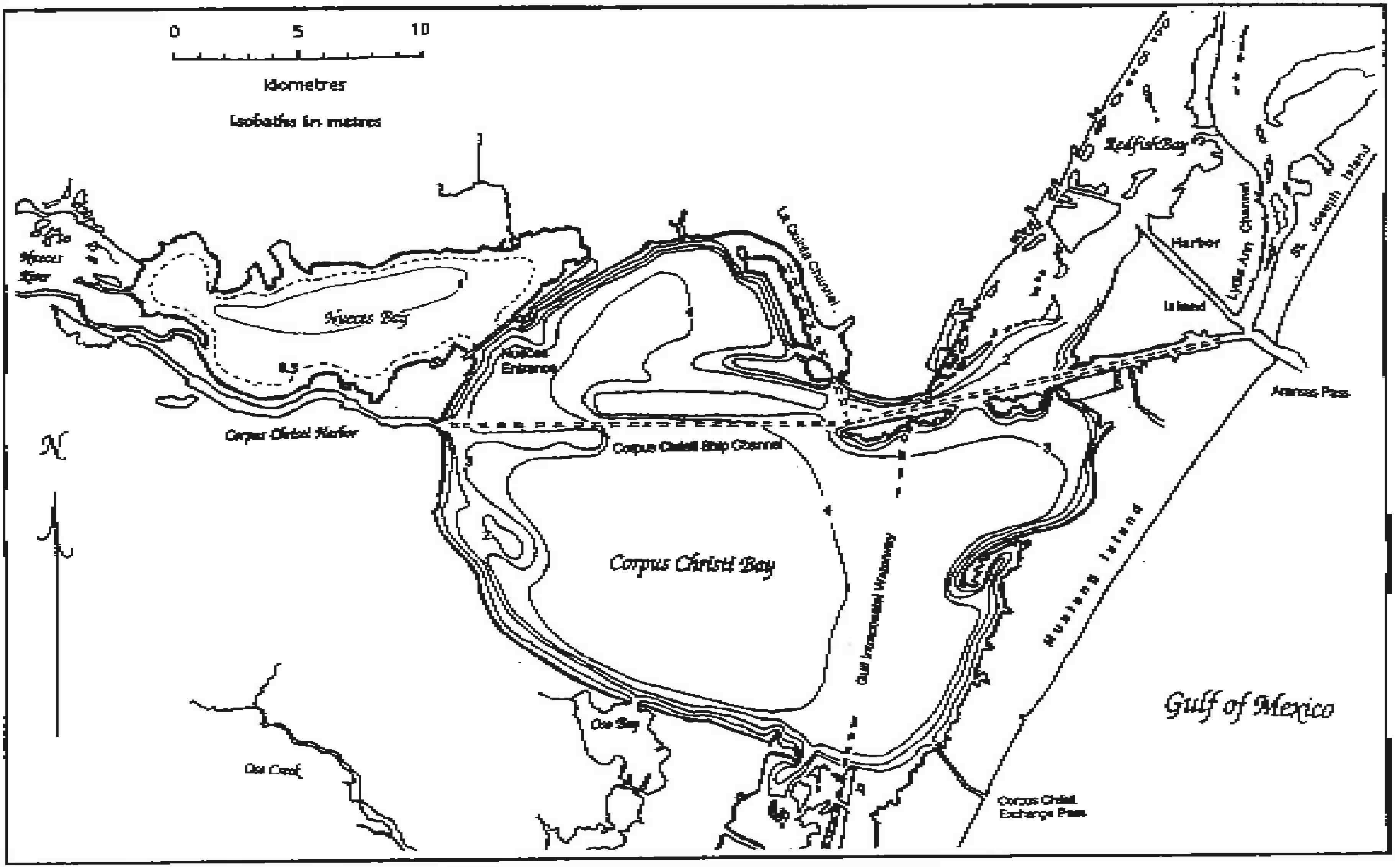


Figure 1-2. Corpus Christi Bay

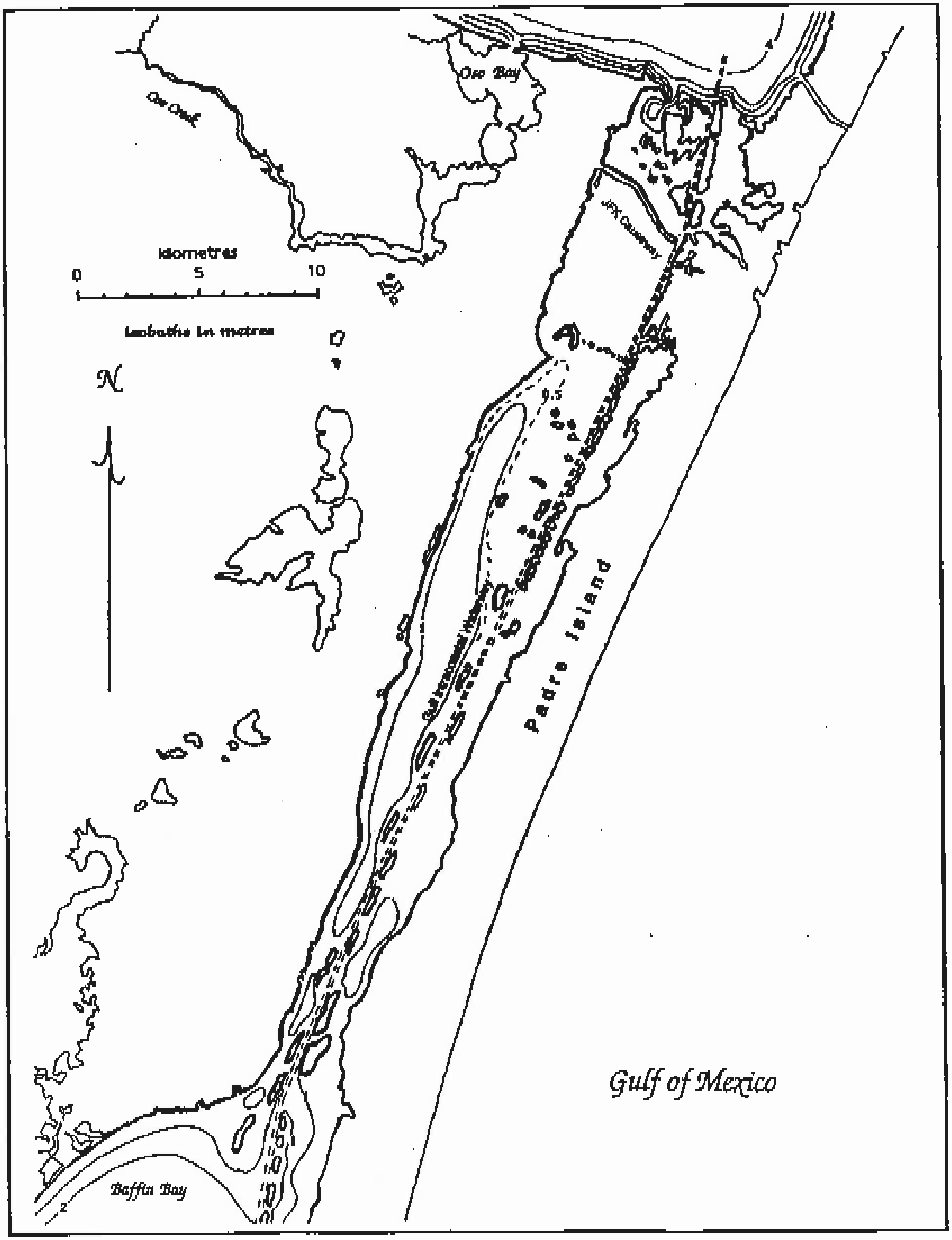


Figure 1-3. Upper Laguna Madre

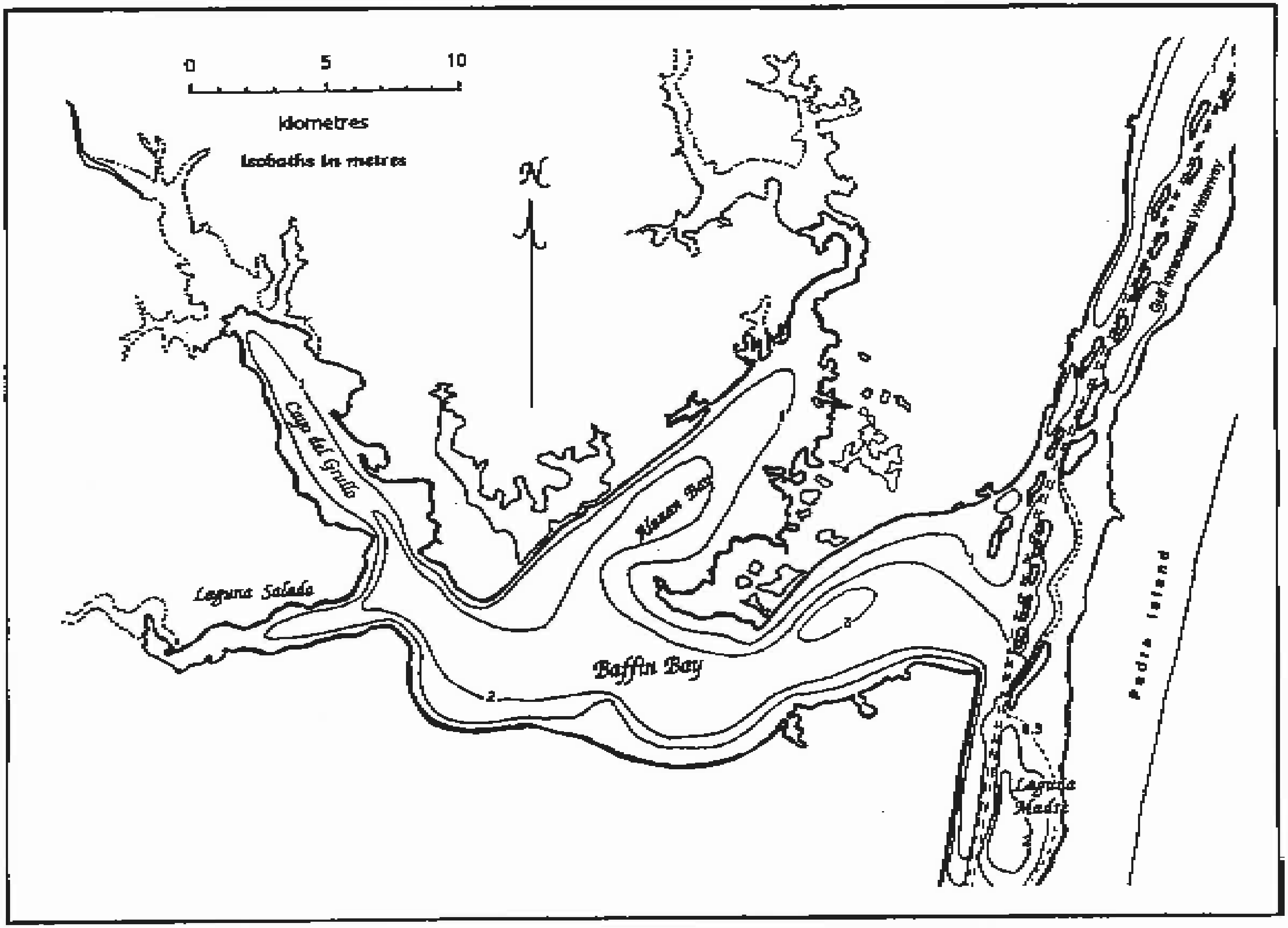


Figure 1-4. Baffin Bay

present status and trends analysis, it was imperative that all data be on hand before we began the analytical aspect of the work; otherwise much of the analysis would have to be repeated. The point in the project calendar at which the data base was declared complete, and analyses began has been referred to as the schedule "hinge point." This poor response directly translated to scheduling problems, and the hinge point was incrementally slid forward in the calendar to accommodate.

Many of the data sets employed in this study exist only in a limited number of hard copies. A major part of the effort of this project was invested in keyboarding this data to create a digital data base. This keyboarding process was delayed by the same problem of poor response, as well as the time necessary in some instances to physically locate the data. The problems of acquiring such data sets would be a formidable obstacle to any future researcher's compiling an adequate data base for Corpus Christi Bay. Therefore, we regard the synthesized digital data base as a major product of the project as it allows future researchers much greater scope in analysis than could be afforded by the data sets normally available to individual scientists.

Procedures of data processing are described in Chapters 3 and 4, the analyzed water and sediment quality data are presented in Chapters 6 and 7 (and appendices), respectively, the possible cause-and-effect processes suggested by associations in the data are discussed in Chapter 9, and a summary of conclusions and list of recommendations are given in Chapter 10. The data base itself is summarized in Chapter 5, along with identified deficiencies and recommendations more directed at monitoring and data management.

Tissue data was included in the Scope of Work for this project. This may have been an afterthought, but it has a logical appeal in that most of the agencies engaged in the collection of tissue chemistry data are also those from which water/sediment chemistry data were sought. This tissue data, however, proved to be more trouble than it was worth. There is very little of it, over a relatively short period of record, and it is reported inconsistently. (This data was more trouble than it was worth at least from the viewpoint of performing a status-and-trends analysis. The data did prove useful to a companion CCBNEP project documenting public health problems, see Jensen et al., 1996.) This part of the data compilation and analytical effort is presented in Chapter 8.

The core of the report is considered to be Chapters 6 and 7. Our philosophy is to present the *facts* of the data in these chapters, reserving the *interpretation* of the data for Chapters 9 and 10. The interpretations postulate conceptual models and may be biased by the predilections of these investigators. Certainly, they will be subject to revision upon additional data collection or more sophisticated analyses. However, the results of Chapters 6 and 7 should stand as *facts*, circumscribed only by the statistical measures employed, and the criteria for rejection or weighting.

2. THE MEASUREMENT OF WATER AND SEDIMENT QUALITY IN CORPUS CHRISTI BAY

The quantification of the quality of water and sediment (and for that matter tissue) in an estuary is accomplished by determination of a suite of chemical and biochemical parameters. Some parameters are routinely measured in the field, such as temp-erature and salinity, but most parameters are determined by laboratory analysis of a water, sediment or tissue sample. Some of these are indicator variables, such as coliforms and BOD, some are constituents which *per se* have major rôles in biochemical processes, such as process compounds, e.g. nitrogen and phosphorus species, or toxic contaminants such as PAH's and pesticides, and some serve in both capacities, such as salinity. This study addressed the following categories of parameters:

- temperature;
- salinity and related parameters;
- suspended sediments, turbidity, and related parameters;
- pH, alkalinity, and related parameters;
- dissolved oxygen and saturation deficit;
- nutrients, *viz.* nitrogen, phosphorus, silica and organic carbon;
- organic contaminants as measured by oil & grease, volatile solids and biochemical oxygen demand;
- chlorophyll-a and phaeophytin;
- coliforms;
- trace metals (total and dissolved);
- trace organics, including pesticides, herbicides, PAH's, PCB's, and priority pollutants

In the present study, data were compiled for 109 water-quality, 83 sediment-quality, and 100 tissue parameters, the more important of which are described below. Complete tabulations of the water quality and sediment quality parameters of this study are given in Table 2-1, and in Tables A-1 and A-2 of the appendix (including sources of data, and uncertainty measures). Some of the more important classes of parameters are discussed briefly in the following sections. Table 2-1 lists the units of measurement employed in this study for each of these parameters, and an abbreviation of up-to-eight characters uniquely identifying the parameter in all of the data presentations in this report, as well as in the digital data files.

This abbreviation is decoded as follows. The first series of 2-3 characters indicates whether the analyte was determined from a water-phase or sediment-phase sample, "WQ" designating the former, and "SED" or "SD" designating the latter. For conventional parameters, the remainder of the abbreviation is a (hopefully) transparent abbreviation for the compound, e.g. WQDO for dissolved oxygen in water, WQFCOLI for fecal coliforms in water, SEDO&G for oil and grease in sediment, etc. For elemental analyses, primarily metals, the compound abbreviation is made up of the prefix "MET" followed by the (1-2 character) chemical abbreviation for the element. In the case of water samples, the sample

Table 2-1
Abbreviations and units for CCBNEP water and sediment parameters

abbreviation	*definition*	*units*
	— water analytes —	
WQALK	total alkalinity (as $CaCO_3$)	mg/L
WQAMMN	ammonia nitrogen	mg/L
WQBOD5	5-day BOD	mg/L
WQCHLA	chlorophyll-a	µg/L
WQCHLB	chlorophyll-b	µg/L
WQCYAN	cyanide	µg/L
WQDO	dissolved oxygen	mg/L
WQFCOLI	fecal coliforms	MPN or colonies/200ml
WQKJLN	total Kjeldahl nitrogen	mg/L
WQNO2N	nitrite nitrogen	mg/L
WQNO3N	nitrate nitrogen	mg/L
WQO&G	oil & grease	mg/L
WQOPD	dissolved orthophosphate (as P)	mg/L
WQOPO4	total orthophosphate (as PO_4)	mg/L
WQORGN	total organic nitrogen	mg/L
WQPH	pH	
WQPHEO	pheophytin-a	µg/L
WQSAL	salinity converted from proxy measures	ppt (‰)
WQSECCHI	Secchi depth of water	m
WQSIO2	dissolved silica (as SIO_2)	mg/L
WQSO4	total sulfate (as SO_4)	mg/L
WQTCOLI	total coliforms	MPN or colonies/200ml
WQTEMP	temperature	degrees C
WQTOC	total organic carbon	mg/L
WQTOTP	total phosphorus (as P)	mg/L
WQTPO4	total phosphate (as PO_4)	mg/L
WQTRANS	transmissivity, over 100 cm path	% (1 m)
WQTSS	total suspended solids	mg/L
WQTURBJ	turbidity of water, JTU	JTU
WQTURBN	turbidity of water, NTU	NTU
WQXTSS	TSS converted from proxy relations	mg/L
WQVOLS	total volatile solids	mg/L
WQVSS	volatile suspended solids	mg/L
WQmetagd	dissolved silver	µg/L
WQmetagt	total silver	µg/L
WQmetasd	dissolved arsenic	µg/L
WQmetast	total arsenic	µg/L
WQmetbt	total boron	µg/L

(continued)

Table 2-1
(continued)

abbreviation	*definition*	*units*
	— water analytes continued—	
WQmetbd	dissolved boron	µg/L
WQmetbad	dissolved barium	µg/L
WQmetbat	total barium	µg/L
WQmetcdd	dissolved cadmium	µg/L
WQmetcdt	total cadmium	µg/L
WQmetcod	dissolved cobalt	µg/L
WQmetcot	total cobalt	µg/L
WQmetcrd	dissolved chromium	µg/L
WQmetcrt	total chromium	µg/L
WQmetcud	dissolved copper	µg/L
WQmetcut	total copper	µg/L
WQmetfed	dissolved iron	µg/L
WQmetfet	total iron	µg/L
WQmethgd	dissolved mercury	µg/L
WQmethgt	total mercury	µg/L
WQmetmnd	dissolved manganese	µg/L
WQmetmnt	total manganese	µg/L
WQmetnid	dissolved nickel	µg/L
WQmetnit	total nickel	µg/L
WQmetpbd	dissolved lead	µg/L
WQmetpbt	total lead	µg/L
WQmetsed	dissolved selenium	µg/L
WQmetset	total selenium	µg/L
WQmetsrd	dissolved strontium	µg/L
WQmetznd	dissolved zinc	µg/L
WQmetznt	total zinc	µg/L
WQ-245T	2,4,5 T	µg/L
WQ-24D	2,4 D	µg/L
WQ-ABHC	alpha-BHC	µg/L
WQ-ACEN	acenapthene	µg/L
WQ-ACENA	acenaphthylene	µg/L
WQ-ALDR	Aldrin	µg/L
WQ-ANTHR	anthracene	µg/L
WQ-BNZA	benzo(a)pyrene	µg/L
WQ-BNZE	benzo(e)pyrene	µg/L
WQ-BNZAA	benzo(a)anthracene	µg/L
WQ-BNZB	benzo(b) fluoranthene	µg/L
WQ-BNZGP	benzo(ghi)perylene	µg/L
WQ-BNZK	benzo(k) fluoranthene	µg/L

(continued)

Table 2-1
(continued)

abbreviation	*definition*	*units*
	— water analytes continued—	
WQ-CHLR	total Chlordane	µg/L
WQ-CHLRC	Chlordane cis isomer	µg/L
WQ-CHRYS	chrysene	µg/L
WQ-DBANE	dibenz(a,h)anthracene	µg/L
WQ-DDD	total DDD	µg/L
WQ-DDE	total DDE	µg/L
WQ-DDT	total DDT	µg/L
WQ-DIAZ	Diazinon	µg/L
WQ-DIEL	Dieldrin	µg/L
WQ-ENDO	Endosulfan I	µg/L
WQ-ENDR	Endrin	µg/L
WQ-FLRA	fluoranthene	µg/L
WQ-FLRN	fluorene	µg/L
WQ-HEPT	heptachlor	µg/L
WQ-HEPX	heptachlor epoxide	µg/L
WQ-HEXA	hexachlorabenzene	µg/L
WQ-I123P	indeno(1,2,3-cd)pyrene	µg/L
WQ-LIND	Lindane (gamma-BHC)	µg/L
WQ-MALA	Malathion	µg/L
WQ-MTHP	methyl parathion	µg/L
WQ-MTHX	methoxychlor	µg/L
WQ-NAPT	napthalene	µg/L
WQ-PAH	total PAH's	µg/L
WQ-PARA	Parathion	µg/L
WQ-PCB	total PCB's	µg/L
WQ-PCP	pentachlorophenol	µg/L
WQ-ODDT	o,p'-DDT	µg/L
WQ-PDDD	p,p'-DDD	µg/L
WQ-PDDE	p,p'-DDE	µg/L
WQ-PDDT	p,p'-DDT	µg/L
WQ-PHNAN	phenanthrene	µg/L
WQ-PYRN	pyrene	µg/L
WQ-SLVX	Silvex	µg/L
WQ-TOXA	Toxaphene	µg/L
WQ-XDDT	Total DDT converted from proxy relations	µg/L

(continued)

Table 2-1
(continued)

abbreviation	*definition*	*units*
	— sediment analytes (dry weight)—	
sedcyan	cyanide	mg/kg
sedkjln	total Kjeldahl nitrogen	mg/kg
sedo&g	oil & grease	mg/kg
sedammn	ammonia nitrogen	mg/kg
sedorgn	total organic nitrogen	mg/kg
sedtoc	total organic carbon	g/kg
SEDtotp	total phosphorus (as P)	mg/kg
sedvols	volatile solids (loss on ingnition)	mg/kg
sedmetag	silver	mg/kg
sedmetal	aluminum	mg/kg
sedmetas	arsenic	mg/kg
sedmetb	boron	mg/kg
sedmetba	barium	mg/kg
sedmetcd	cadmium	mg/kg
sedmetco	cobalt	mg/kg
sedmetcr	chromium	mg/kg
sedmetcu	copper	mg/kg
sedmetfe	iron	mg/kg
sedmethg	mercury	mg/kg
sedmetmn	manganese	mg/kg
sedmetni	nickel	mg/kg
sedmetpb	lead	mg/kg
sedmetse	selenium	mg/kg
sedmetsr	strontium	mg/kg
sedmetzn	zinc	mg/kg
sed-245t	2,4,5 T	µg/kg
sed-24d	2,4 D	µg/kg
sed-abhc	alpha-BHC	µg/kg
sed-acen	acenapthene	µg/kg
sed-acyn	acenaphthylene	µg/kg
sed-aldr	Aldrin	µg/kg
sed-anth	anthracene	µg/kg
sed-bnza	benzo(a)pyrene	µg/kg
sed-bnze	benzo(e)pyrene	µg/kg
SD-bnzaa	benzo(a)anthracene	µg/kg
SD-bnzb	benzo(b) fluoranthene	µg/kg
SD-bnzk	benzo(k) fluoranthene	µg/kg
SD-bnzgp	benzo(ghi)perylene	µg/kg

(continued)

Table 2-1
(continued)

abbreviation	*definition*	*units*
	— sediment analytes continued—	
sed-chlr	total Chlordane	µg/kg
sd-chlrc	Chlordane cis isomer	µg/kg
sed-chry	chrysene	µg/kg
sed-ddd	total DDD	µg/kg
sed-dde	total DDE	µg/kg
sed-ddt	total DDT	µg/kg
sed-diaz	Diazinon	µg/kg
SD-dbane	dibenz(a,h)anthracene	µg/kg
sed-diel	Dieldrin	µg/kg
sed-endo	Endosulfan I	µg/kg
sed-endr	Endrin	µg/kg
sed-flra	fluoranthene	µg/kg
SD-flrn	fluorene	µg/kg
sed-hept	heptachloride	µg/kg
sed-hepx	heptachloride epoxide	µg/kg
sed-hexa	hexachlorobenzene	µg/kg
SD-I123p	indeno(1,2,3-cd)pyrene	µg/kg
sed-lind	Lindane (gamma-BHC)	µg/kg
sed-mala	Malathion	µg/kg
sed-mthp	methyl parathion	µg/kg
sed-mthx	methoxychlor	µg/kg
sed-napt	napthalene	µg/kg
sed-pah	total PAH's	µg/kg
sed-para	Parathion	µg/kg
sed-pcb	total PCB's	µg/kg
sed-pcp	pentachlorophenol	µg/kg
sed-pddd	p,p'-DDD	µg/kg
sed-pdde	p,p'-DDE	µg/kg
sed-pddt	p,p'-DDT	µg/kg
sed-oddt	o,p'-DDT	µg/kg
sed-oddd	o,p'-DDD	µg/kg
sed-odde	o,p'-DDE	µg/kg
sed-pery	perylene	µg/kg
SD-phnan	phenanthrene	µg/kg
SD-pyrn	pyrene	µg/kg
SED-slvx	Silvex	µg/kg
sed-toxa	Toxaphene	µg/kg
sed-tbt	tributyltin	µg/kg
sed-xddt	DDT converted from proxy relations	µg/kg

may have been filtered, in which case the analysis is presumed to represent the dissolved metal; or the sample may not have been filtered, in which case the analysis is presumed to represent both the dissolved and suspended portions of the metal. The former is indicated by the letter "D" for "dissolved", and the latter by the letter "T" for "total." For example, WQMETASD refers to the arsenic in a filtered water sample, WQMETSET to the selenium in an unfiltered water sample, and SEDMETPB to lead in a sediment sample. Finally, all volatile organics are flagged by a hyphen in the abbreviation after the water/sediment phase designation. For example, WQ-ACEN refers to acenapthene in a water sample, and SED-LIND to the lindane in a sediment sample.

There are several classes of parameters that measure (or can be interpreted to measure) the same essential property. For example, salinity can be estimated from measurements of chlorides concentration, total dissolved solids, density, conductivity, and light refraction. Different data collection programs in Corpus Christi Bay may employ different measures, depending upon objective, convenience and tradition. The relations between parameters are considered here, for two purposes. First, from an analytical viewpoint, the use of one parameter may have conceptual advantages over another, e.g. DO deficit may be more indicative of oxygen conditions than the concentration of dissolved oxygen itself. Second, while related parameters are technically distinct, the fact that they can be associated and may be converted from one to another means that a much denser and longer-duration data set can be compiled by converting these to a common parameter. These are referred to as "proxy" relationships. For a few of the variables treated in the present study, proxy relationships proved valuable. These are summarized where appropriate in the following sections.

2.1 Hydrographic indicators

The parameters temperature and salinity are easily measured, and have been routinely determined for some time, therefore the data base is most extensive for these variables. Some of the older methods of determination involve water sampling, but the newer methods can be performed *in situ*. These have particularly benefited from the development of field instrumental techniques, by which the parameters are measured by an electrometric probe with meter readout or remote data logger. These probes have also permitted the determination of vertical profiles of these parameters without the need for water sampling. Rather, the probe can be lowered to the desired depth, and the measurement read from the deck readout.

Temperature and salinity exhibit considerable variability in Corpus Christi Bay, temperature due to the local heat-exchange processes at the surface, and salinity due to watermass movement within the estuary in conjunction with high spatial gradients. After temperature and salinity, pH is probably the most commonly measured parameter, again because of its simplicity. Generally, pH exhibits little variability, due to the high buffering capacity of seawater, but for this reason departures from the range 7-9 are especially significant.

Salinity is one of the quintessential quality elements of estuarine waters, being determined fundamentally by the intermixing of fresh and oceanic waters. As a virtually conservative parameter, easily measured, and ubiquitous, it is an excellent watermass tracer. It is also a key ecological indicator, as it affects the suitability of habitat due to varying osmoregulation capabilities of organisms. Salinity further affects many chemical reactions and sedimentation. Any direct impact on salinity has the potential of indirect consequences for ecosystem structure and function.

Since there are large spatial gradients in salinity and it exhibits high temporal variability, for work in estuaries a lower degree of precision in salinity determination can be accepted than the case either in totally fresh or oceanic systems (Head, 1985, Ward and Montague, 1996). This means that data can be employed from a variety of protocols and parameters. For this data compilation, these various parameters were converted to equivalent salinity.

Salinity originally measured the dissolved solids in seawater, which are dominated by halogen salts. A simpler measure was to determine the salts of a single halogen, *viz.* chlorine, and employ the empirical law of constant proportions (Forchhammer's Law). The relation between salinity and chlorinity based upon early work of Knudsen is approximately (Defant, 1961, Wallace, 1974)

$$S = 0.03 + 1.805 \bullet Cl$$

for S and C in ‰. A century later, this relation was re-evaluated as

$$S = 1.807 \bullet Cl \qquad (2\text{-}1)$$

Certainly to the accuracy necessary for estuarine work, this is a satisfactory means of interconverting.

One of the most common methods of salinity measurement is via conductivity. (In fact, in oceanography, the new practical salinity scale *defines* salinity in terms of conductivity.) Conductivity is a strong function of temperature so the temperature at which the measurement applies is essential. Generally, a reported conductivity will be either at ambient temperature, i.e., the temperature of the water when the measurement was taken, or compensated to a standard reference temperature of 25°C. (Modern inexpensive conductivity meters perform this compensation internally.) For a fixed temperature, conductivity varies nearly linearly with salinity. A regression based upon the data of USNHO (1956) is

$$\begin{aligned} S &= 0.000588 \bullet C & &\text{for } C < 17{,}000\ \mu\text{mhos} \\ S &= 0.000679 \bullet C - 1.543 & &\text{for } C > 17{,}000\ \mu\text{mhos} \end{aligned} \qquad (2\text{-}2)$$

where C is conductivity at 25°C and S is salinity in ‰. All of the conductivity data from Corpus Christi Bay were either compensated to 25°C, or have been converted to salinity by the collecting agency, so the further complication of correcting for ambient temperature was not necessary. (Ward and Armstrong, 1992a, present an equation appropriate for estuaries.) If chlorinity rather than salinity is desired, the above relations must be combined to relate conductivity and chlorinity, *viz*.

$$C = \begin{cases} 2661 \cdot Cl + 2272 & \text{for } Cl > 5.5\ \text{‰} \\ 3073 \cdot Cl & \text{for } Cl < 5.5\ \text{‰} \end{cases} \qquad (2\text{-}3)$$

Density is a fundamental physical property of the estuary water. It varies as a function of temperature and salinity. Some determinations of salinity, especially in the early years, are performed by measuring density with precision hydrometers. Again, the oceanic relation is basic. The equation of state for seawater is empirical, and has most recently (UNESCO, 1981) been expressed as a best-fit multinomial with 15 coefficients. For present purposes, we retain only the higher-order terms, to obtain the approximate relation:

$$\rho = (a_4 + b_4 T + a_5 T^2)\, S + (a_1 + b_1 T + a_2 T^2) \qquad (2\text{-}4)$$

where

$$a_1 = 999.8426 \qquad b_1 = 6.794 \times 10^{-2}$$
$$a_2 = -9.0953 \times 10^{-3} \qquad a_4 = 8.245 \times 10^{-1}$$
$$a_5 = 7.644 \times 10^{-5} \qquad b_4 = -4.090 \times 10^{-3}$$

for salinity S in parts per thousand, temperature T in degrees Celsius, and density ρ in kg/m^3. This approximation is more than adequate for the accuracy necessary in estuary work (see Ward and Armstrong, 1992a, who present numerical comparisons between the complete equation of state and that given by the above equation).

One additional field measure of salinity is the refractive index of water. The field instrument used for this purpose is a portable refractometer that is calibrated for a direct read-out of salinity (the Goldberg refractometer). For present purposes, therefore, no units conversion is necessary, but in the data processing procedures of Chapter 4 note is made of when this methodology is employed, for establishing a level of uncertainty in the data. Behrens (1965) carried out an extensive evaluation of the accuracy of the Goldberg refractometer and determined the error to be ±1‰ for salinities < 70‰.

Generally, in the field data from the Corpus Christi Bay system, only one of these measures is employed for determination of salinity, so the only decision available in analyzing the data is the proper conversion. On occasion, there is a choice, which provides an opportunity to cross-compare different measures. In the TNRCC Statewide Monitoring Network data base, both field and laboratory conductivity measurements may be available for a given sample, and occasionally there may be a laboratory determination of chlorides for the same sample as well. Where all three variables were measured, we can test the internal consistency of the data. This was carried out by Ward and Armstrong (1992a, 1992b) in the

Galveston Bay NEP data compilation. For Corpus Christi Bay, this same analytical approach was repeated, and the same problems are evident. However, in this study we gained additional insight into the sources of the discrepancies. This is summarized below; much more detail is given in the companion data-base report, Ward and Armstrong (1997a).

A clear manifestation of a problem is wide scatter in the field versus laboratory measurements of conductivity, as shown in Fig. 2-1. Considering that conductivity probes are among the simplest instruments to maintain and employ, and granting that some noise may be expected due to the hostile conditions of field measurements, this scatter is excessive, and does not engender a feeling of warmth and comfort in the overall quality of the data base. After study of that subset of over 1500 measurements for which simultaneous values for chlorides and both conductivities are given, our conclusion is that most of the "noise" is contributed by the laboratory conductivities originating at the Texas Department of Health. Part of this widespread discrepancy is due simply to inferior accuracy in the laboratory determinations, which were subjected to extreme dilution to bring the conductivity into the narrow, low range of the laboratory meter, then the measured value was scaled back up by the reciprocal of the dilution. This introduces several potential sources of error as delineated in Ward and Armstrong (1997a). Though field conductivity meters are presumably less accurate than low-range laboratory meters, they measure conductivity directly without any necessity for dilution.

Another contributor to the scatter is due to the fact that a significant proportion of the reported laboratory values are not really measurements, but are "invented data," i.e. entries Based On Graphical or Arithmetical Suppositions, BOGAS data. The BOGAS data were discovered in scatterplots of lab conductivity versus either field conductivities or chlorides, in which there appeared two distinct regressions, of which one corresponded to the theoretical relation, equation (2-3). This is particularly evident in the plot of lab conductivity versus chlorides, in which some of the data fall along a line with almost *zero* scatter, see Fig. 2-2. This line proved to be the relation $y = x \cdot 4 = x \cdot 8/2$. The theoretical relation between conductivity and chlorides is given approximately by $y = x \cdot 5/2$. We infer that for much of the data set, only one of chlorides or conductivity was actually measured, and the other was computed based upon the (correct) rule of thumb that conductivity is 5/2 times chlorides, and upon the (incorrect) rule that conductivity is 8/2 times chlorides. A comparison of the *field* conductivities and the chlorides shows a well-behaved relation that centers upon the theoretical oceanographic relation, with realistic scatter, Fig. 2-3.

We conclude that the lab *chlorides* and *field conductivity* values are real and the *lab conductivities* are (potentially) BOGAS. We note that this practice of supplying BOGAS data, apart from corrupting the data base with non-measured values, offers one more degree of freedom for human error, and indeed this is almost certainly the reason for the second spurious regression in Fig. 2-2. In this compilation, we expunged the BOGAS data from the data base by observing the priority of acceptance of salinity measures: chlorides, followed by field conductivity, followed by lab conductivity. This same priority was employed for

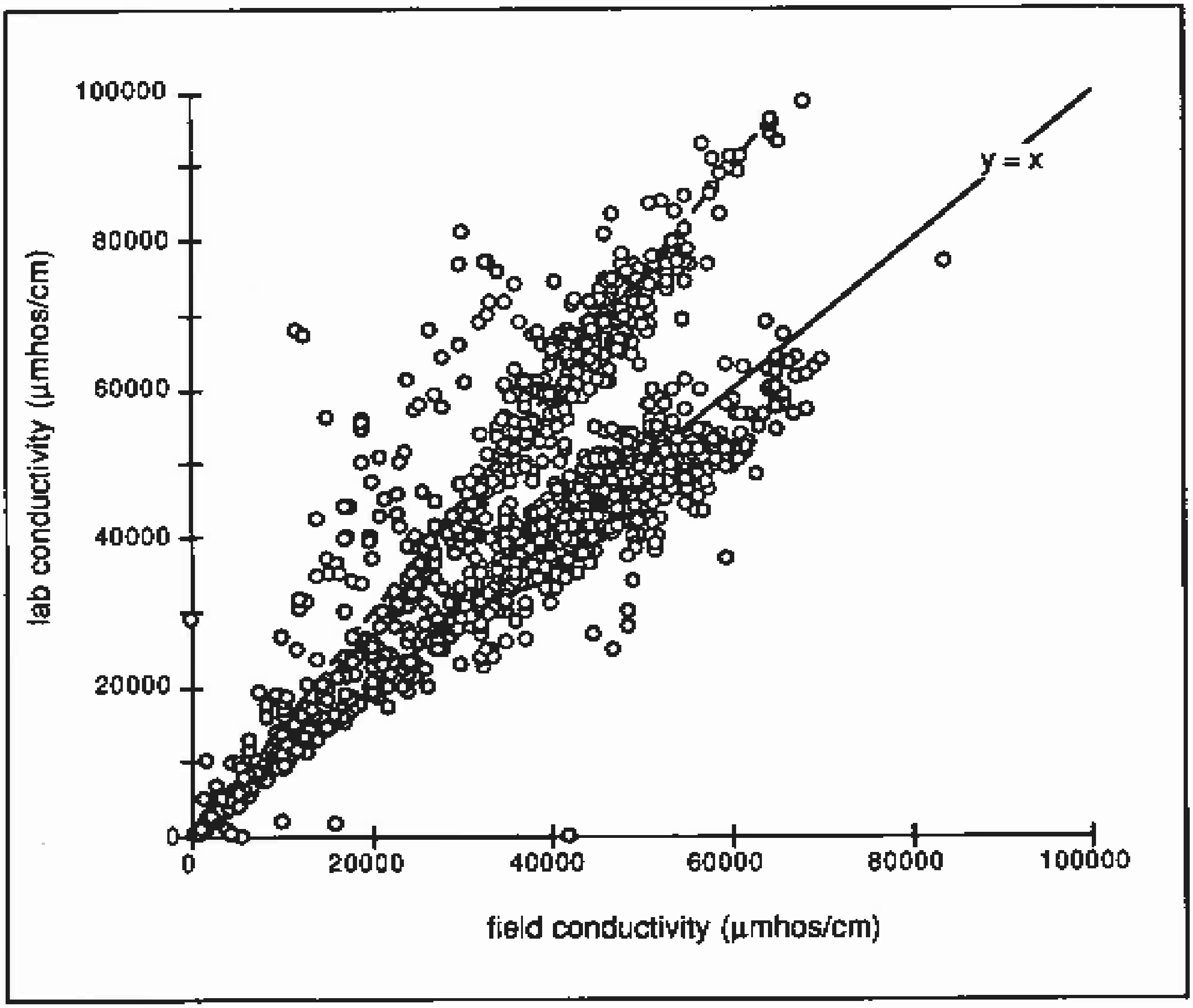


Figure 2-1. Scatterplot of field and laboratory conductivities from SMN data base for Corpus Christi Bay system

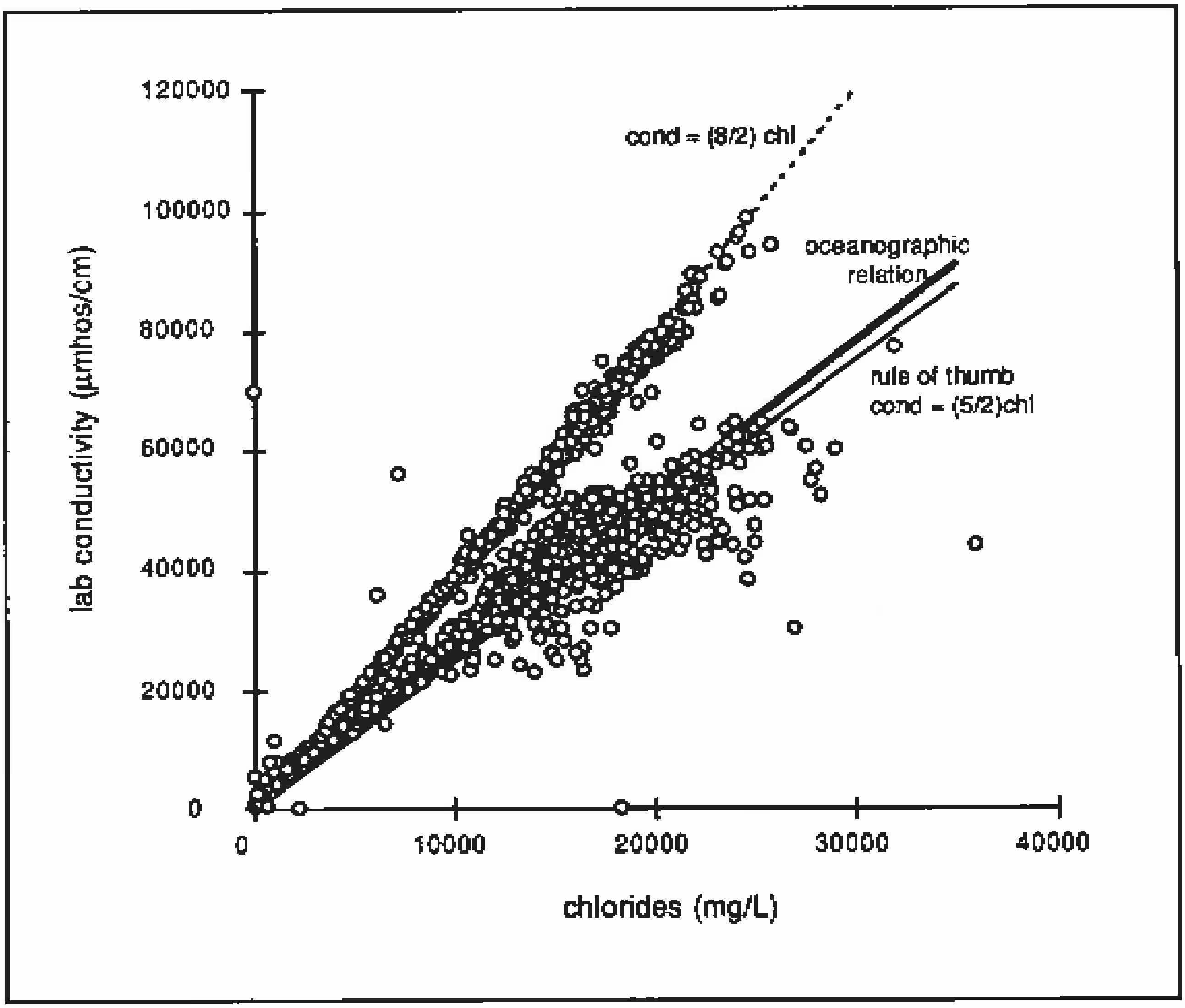


Figure 2-2. Scatterplot of lab conductivity versus chlorides from SMN data base for Corpus Christi Bay system.

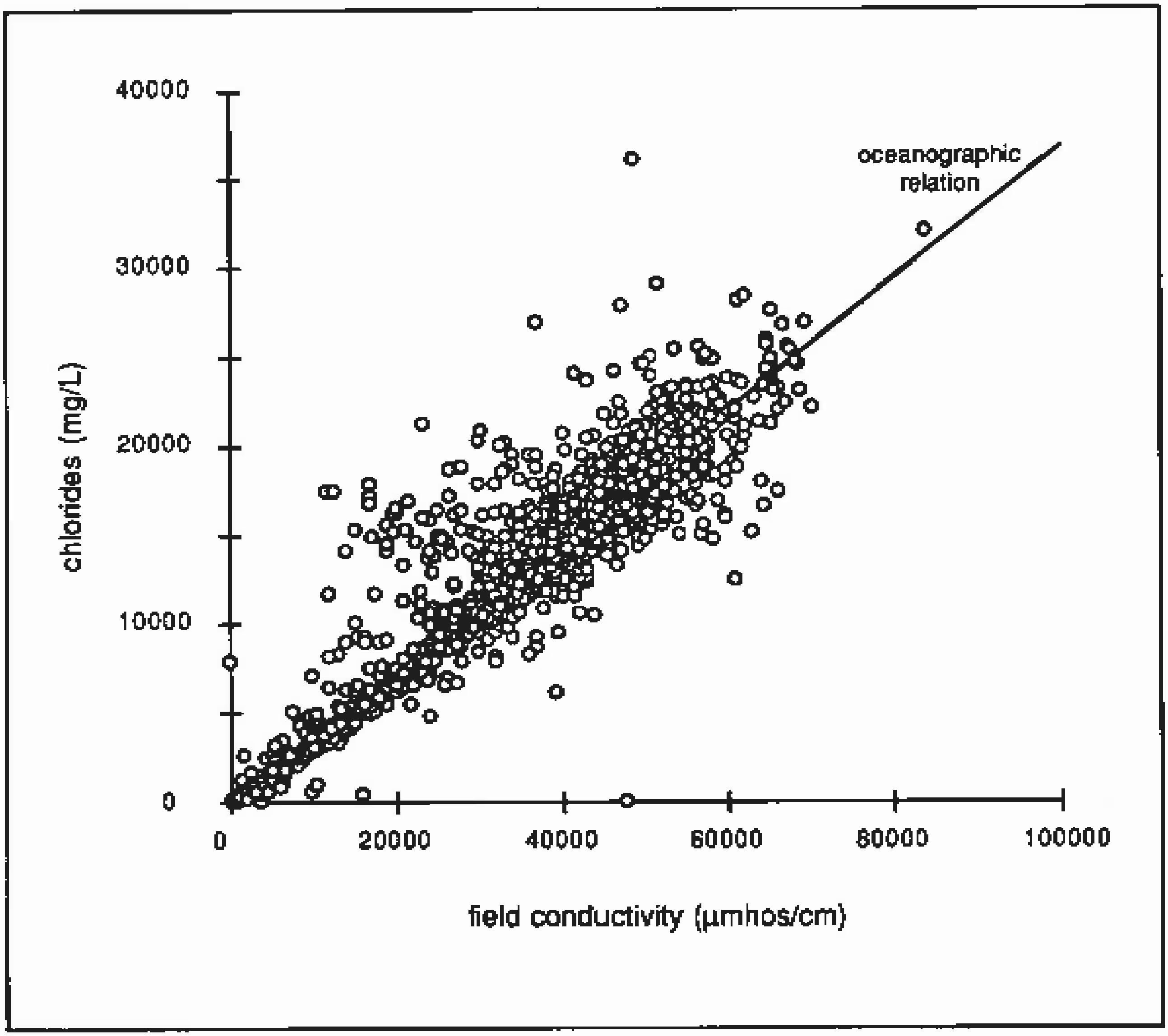


Figure 2-3. Scatterplot of chlorides versus field conductivity from SMN data base for Corpus Christi Bay system.

the older data in both the TWDB Coastal Data System files, and the TDH Seafood Safety data base, for which the TSDH laboratory performed the chlorides and lab conductivity analyses. (The practice of inserting BOGAS data is addressed further in Section 5.3.2.)

Among the data from the Marine Science Institute are field salinities determined both by conductivity and by hand-held refractometer, offering an opportunity for comparison. A scatterplot of those data from the Corpus Christi Bay synoptic surveys is shown in Fig. 2-4. The linear correlation coefficient is 0.825, not particularly high in view of the common assumption that refractometry is a suitable substitute for more precise methods (even for the relatively low precision demands of estuary work). The standard error of the estimate is about 3 ‰, independent of whether the regression is constrained through (0,0). Presuming that the determination by conductivity is the more precise measurement, this standard error would then correspond to the estimated accuracy of the refractometer. It is interesting to compare this to the data of Behrens (1965) who found a standard error of approximately ± 1‰ for the range of salinities represented here, approximately three times better than indicated in the data of Fig. 2-4. His study was carefully executed, and his data represent probably the very best precision that the refractometer is capable of, while the data analyzed in Fig. 2-4 are more typical of the usual field operation of the refractometer.

2.2 Dissolved oxygen

DO is one of the fundamental indicators of aquatic health, since it determines the ability of aerobic organisms to survive. With the development of electrometric probes for DO—a welcome technology for anyone who has ever performed Winklers in a rocking boat—field measurements of DO have increased geometrically, and are now a routine component of most *in situ* monitoring. The data base for DO is therefore approaching that for temperature and salinity, especially in the last two decades, though the data from the 1950's to the 1960's are principally laboratory determinations on water samples.

DO is introduced into the water column principally through "reaeration," the mechanical process of surface transport from the atmosphere, and through photosynthesis. Therefore DO can serve as an indicator of both mechanical aeration and the intensity of primary production. The primary depletion of DO is due to biochemical stabilization of organics through the respiratory process of the biological community (see Section 2.4 below), and low DO's are traditionally linked to the presence of oxygen-demanding pollutants and/or very high rates of primary production.

One of the key controls on the concentration of DO is its solubility, which is a strong function of temperature and salinity. APHA (1985) compiled data on oxygen solubility and offered a nonlinear regression equation with eight coefficients for direct calculation, see also Head (1985) and references therein. This expression is much more precise than is required here, even if the

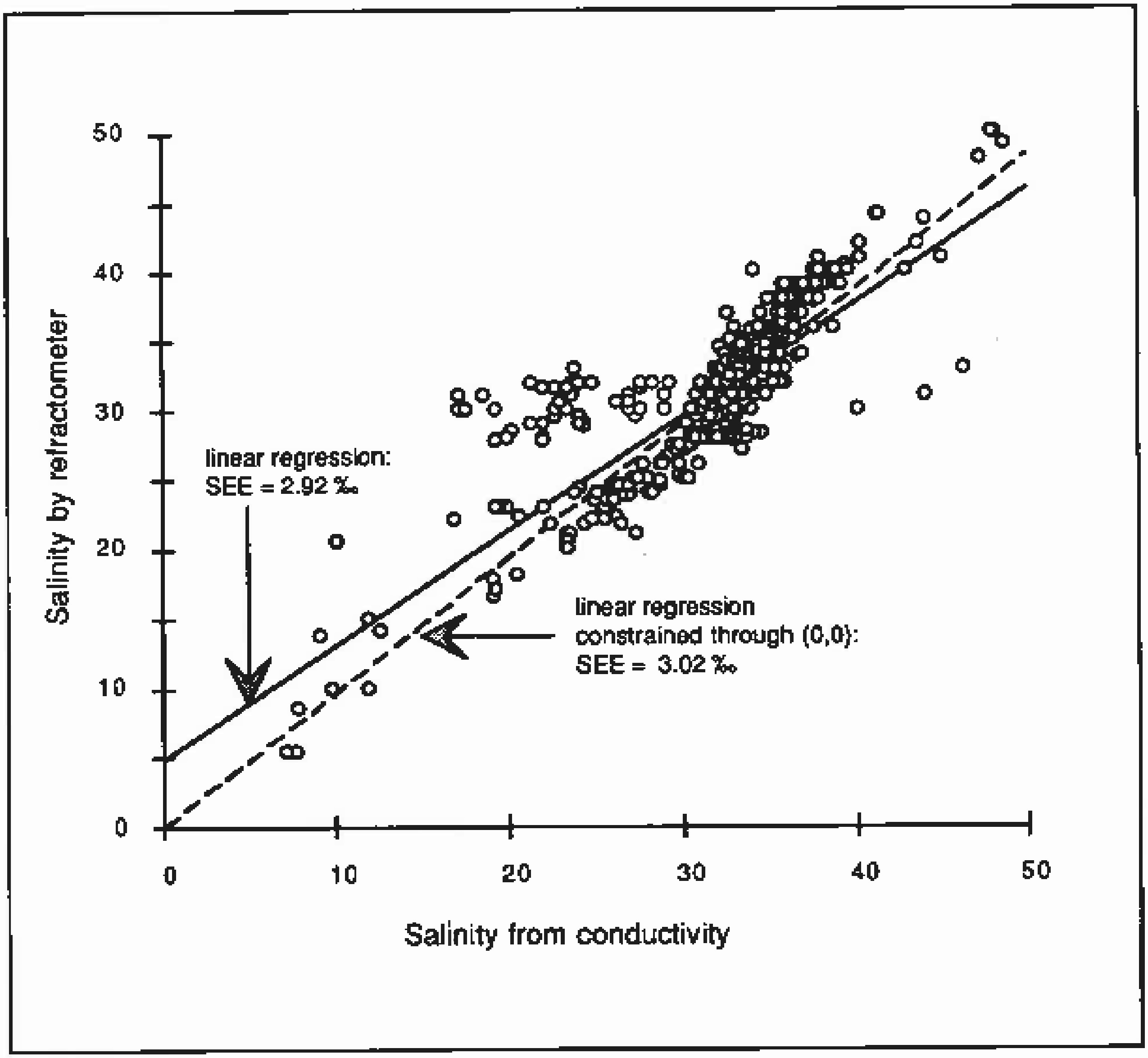


Figure 2-4. Scatterplot of companion salinities determined from conductivity and from refractometer, from data of University of Texas Marine Science Institute (Project 17)

coefficients were correct (which they are not). Ward and Armstrong (1992a) sought a simpler functional equation, finally selecting the expression:

$$C_S = 5\,(100 - Cl)\,/\,(T + 35) \tag{2-5}$$

as a compromise between simplicity and accuracy (see Fair and Geyer, 1954), being accurate to better than 0.1 mg/L. Here C_S denotes DO saturation in mg/L, Cl is chlorinity in parts per thousand and T is temperature in degrees Celsius. The coefficients were re-evaluated by Ward and Armstrong (1992a) using the data in APHA (1985), who presented tabular comparisons to demonstrate the general accuracy of this relation for the normal estuarine range of chlorinity and temperatures.

The functional dependence of solubility on temperature and salinity given by (2-5) illustrates that saturation—and hence DO concentration—will vary substantially over the year, see Fig. 2-5. As temperatures range from perhaps 5° to 35°C and chlorinity from 0 to in excess of 45 ‰, and if the cold temperature is correlated with low salinity and the highest temperature with highest salinity, a not-unreasonable assumption for Corpus Christi Bay, the total excursion in solubility is from 5 to 14 mg/L. This high range of natural variability can mask variations in DO of importance in diagnosing water-quality problems. Accordingly, two associated parameters of dissolved oxygen are defined: the oxygen *deficit*

$$D = C_S - C \tag{2-6}$$

and fraction of *saturation*

$$Sat = 100\,C\,/\,C_S \tag{2-7}$$

where C is DO concentration and D is DO deficit, both in mg/L, and *Sat* is saturation in per cent. The use of these parameters effectively removes the influence of varying temperature and salinity, and allows a more direct interpretation of the (transformed) DO measurements in terms of water quality.

Interpretation of the DO "climate" requires any two of these three parameters. For present purposes, we employ the DO concentration and the DO deficit. The more important of these is the deficit, because it can be shown that if DO kinetics are first-order (or, more generally, are approximately linear in DO concentration) then the total temperature and salinity variation is absorbed in the solubility: the corresponding DO deficit has no temperature/salinity dependency. Deficit should therefore yield more meaningful statistics and trends, because the "noise" contributed by variations in temperature and salinity is eliminated. Deficit by itself, however, cannot be interpreted biologically: a deficit of a given magnitude, 4 mg/L for example, may be biologically limiting in summer and biologically unimportant in winter.

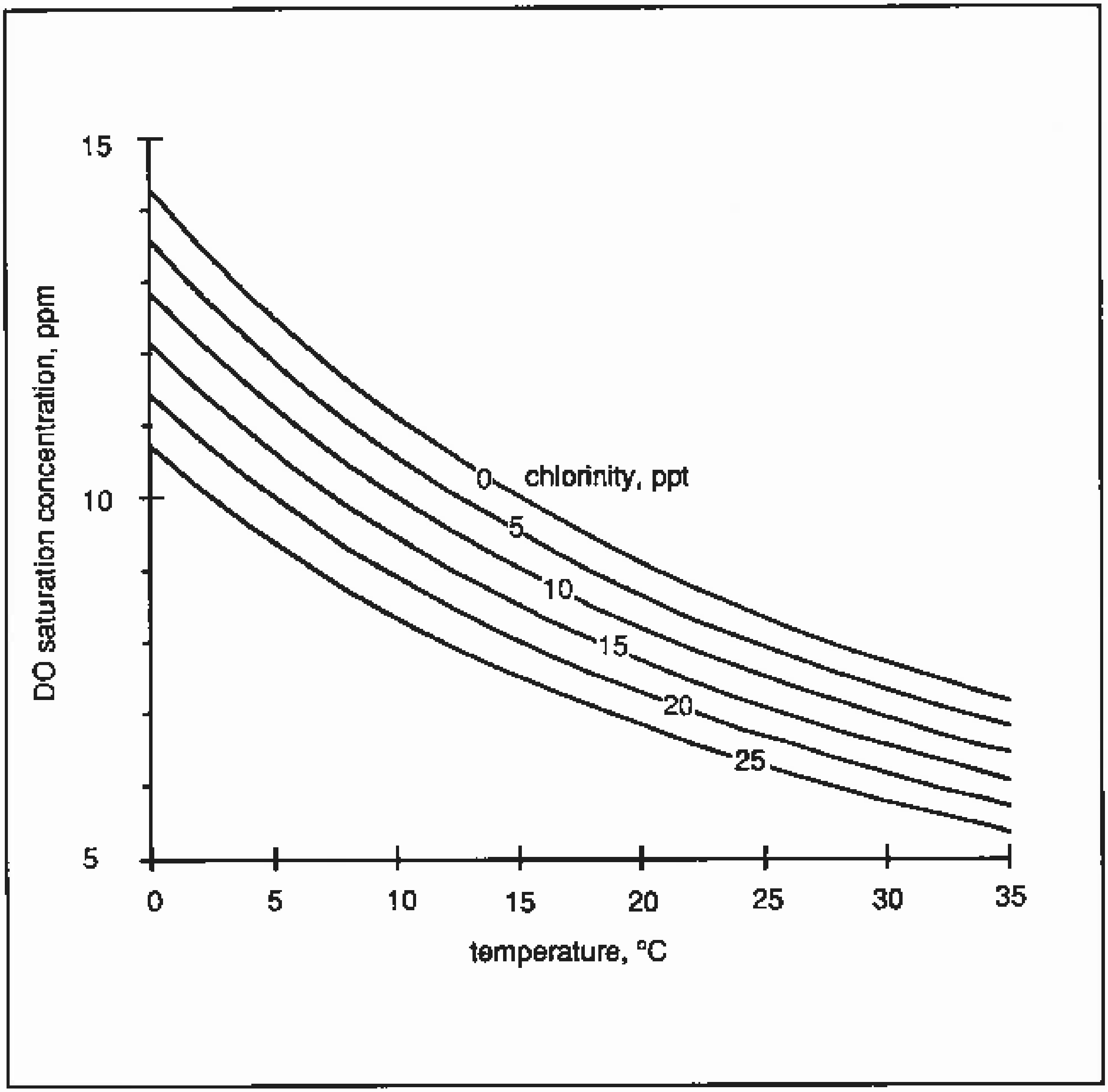


Figure 2-5. DO solubility for ranges of chlorinity and temperature typical of Corpus Christi Bay

2.3 Suspended solids

Solid particles of greater density than water but which are small enough to be carried, perhaps briefly, in suspension by fluid flow are referred to as suspended matter/sediment/solids/particulates or, if clear in context, the adjective "suspended" may be dropped. Suspended solids are traditionally measured by a simple filtration (0.45 microns). Ward and Montague (1996) comment that, unlike the situation in freshwater systems, suspended sediment "is a normal, ubiquitous component of the estuarine environment." This observation cannot be overemphasized, and is especially true for Corpus Christi Bay. Sediments enter the water column by mobilization from the bed, transport into the estuary by freshwater inflows, discharge of waste streams, transport from the nearshore littoral zone by tidal currents, by erosion of the shoreline, and deposition from the aeolian flux in the lower atmosphere.

Suspended solids play an important rôle in determining light penetration and primary productivity in the water column. A closely related property therefore is turbidity, which refers to the interference with the passage of light by suspended matter in the water, and is an indirect indicator of the concentration of such suspended matter. Further, there are methods of making turbidity-related observations in the field. While turbidity has value in itself as a water-quality indicator, our present interest is in its use as a proxy measure of suspended solids. The following review was presented by Ward and Armstrong (1992a).

The traditional method of viewing a candle flame through a vertical tube containing the water sample motivated the definition of the Jackson Turbidity Unit (JTU), see APHA (1985). Modern electrometric optics offer an alternative to the traditional Jackson turbidimeter (e.g., APHA, 1985, Lamont, 1981, Kirk, 1983). The reduction in transmitted light intensity, as measured by a transmissiometer, is expressed as a fraction of the source intensity (per cent transmittance) or in terms of an extinction coefficient. A few measurements of this type exist from Corpus Christi Bay, primarily carried out by the Southwest Research Institute field program. Alternatively, nephelometers measure light scattering at 90° and the measurement is reported in Nephelometric Turbidity Units, which are defined to be numerically about the same as JTU's. Laboratory turbidity measures are calibrated by standard silica suspensions, so as to eliminate the source of variation due to suspended particles of different constituency and geometry. Unfortunately, the numerical equivalence of JTU's and NTU's holds only for the calibration compound. For different types and distributions of suspended matter, NTU's and JTU's depart. Further, each is an index and does not *per se* correspond to a physical property of the water. When the reference suspension in the nephelometric procedure is the formazin polymer, the results are often reported as FTU; for present purposes, we regard these as equivalent to NTU.

The depth of the Secchi disc has for many years been the limnologist's and oceanographer's standard means for field measurement of turbidity (Hutchinson, 1957). Unfortunately, the relation between Secchi depth and conventional measures of turbidity is murky at best. Hutchinson (1957) comments, "we should not expect to find more than a rough correspondence between the transmission and Secchi disk transparency of a series of lakes," on the basis of a brief analysis of the differing responses of the transmissiometer and the Secchi disk. A deeper analysis offers somewhat more optimism; Preisendorfer (1986) is the last word on the subject. Effler (1988) combined literature optical theory with field measurements from a number of lakes to arrive at the relation

$$SD = N / [\sqrt{(a^2 + 0.256\,ab)} + (a + b)]$$

where a and b are the absorption and scattering coefficients, resp., in dimensions $[L^{-1}]$. Here N is a constant, probably a weak function of other optical properties including spectral distribution of light, in the range 8.0-9.6 . Since b>>a usually, for nephelometric turbidity T (roughly proportional to b, with a constant $\approx 1 \pm 25\%$ for b in m^{-1} and T in NTU) the approximate relation becomes:

$$SD = N'' / T \qquad (2\text{-}8)$$

where N" ranges about 5-10 for SD in meters and T in NTU's, depending on other optical properties of the water. Vis-a-vis application of the relation (2-8) to the turbid waters of a shallow estuary, in contradistinction to the lakes addressed by Effler (1988), the water is muddied if the water is muddy, because SD becomes decreasingly sensitive to T as T becomes large. (Holmes, 1970, developed an inverse exponential relation between SD and an optical parameter related to an extinction coefficient, which has the same asymptotic behavior.)

Relating turbidity to suspended solids is even more opaque. From Mie theory, a relation would be anticipated between suspended particles and the scattering coefficient b of the form

$$b = \pi \Sigma n_i r_i^2 / 4$$

where n_i is number of particles of mean radius r_i per unit volume. This implies b = A•SS for SS the suspended solids concentration. (Actually, b may be taken as the total extinction coefficient.) From British coastal waters A lies in the range 0.25-0.50 for SS in mg/L (Jones and Willis, 1956). Di Toro (1978) found A=0.40 for San Francisco Bay. Since T is proportional to b (some authors assert $T \approx b \pm 25\%$), we have T proportional to SS.

In summary, there is good reason to expect an inverse relation between suspended solids and Secchi depth, SS = B/SD, and a direct proportional relation between suspended solids and turbidity, SS = A T , with constants (from the above literature values) on the order of $10<B<50$ and $0.2<A<0.5$, for SD in meters, SS in mg/L, and T in JTU or NTU. The problem now is to test these relations against data from Corpus Christi Bay to verify the functional form, and to determine the appropriate values of the constants.

Virtually the only extensive paired measurements of Secchi depth, turbidity and TSS are those of the TNRCC SMN data base. The SMN data of paired turbidity and TSS measurements are shown in Fig. 2-6. That each is a noisy measurement

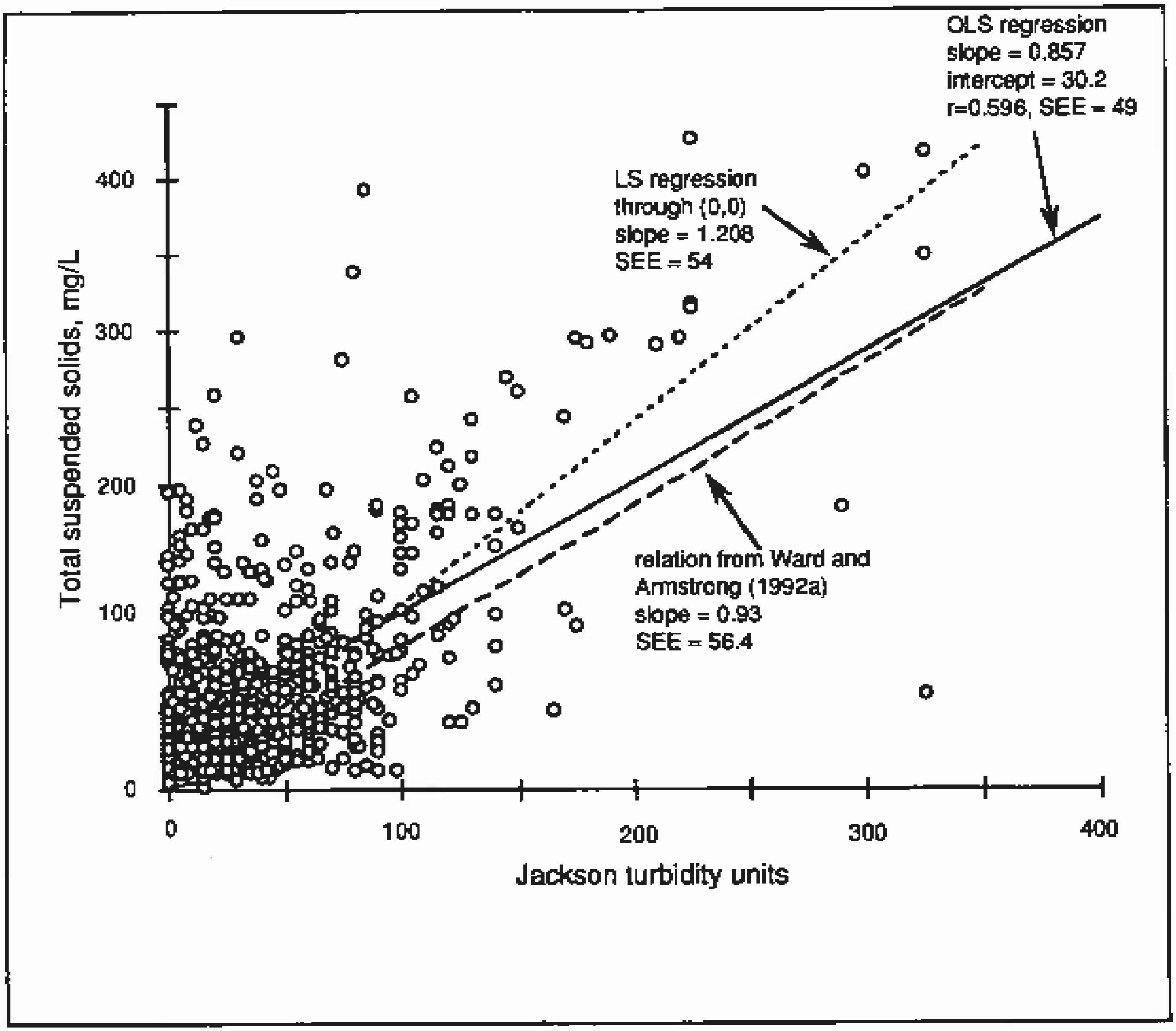


Figure 2-6. TSS versus turbidity in Corpus Christi Bay

is apparent. However, the data (856 data points) of Fig. 2-6 are consistent with the hypothesis of a proportional relation. The zero-intercept least-squares (LS) line has a standard error about 10% larger than that of the ordinary least-squares (OLS) line and is physically better based. The slope of this zero-intercept line is about 1.2, which is about half the value to be expected from the literature values given above. The same basic relation with a slope of about 0.93 was found by Ward and Armstrong (1992a) from 1350 data points for the bays of the Galveston system. This relation applied to the Corpus Christi data base yields a standard error virtually the same as the LS zero-intercept line. Given the insensitivity of the standard error to the slope of the regression, the larger data base for Galveston Bay from which the Ward and Armstrong (1992a) relation was derived, and the fact that this relation is equally consistent with the Corpus Christi data, we have elected to employ it as a proxy relation giving TSS as a function of JTU's.

Figure 2-7 addresses the inverse relation between TSS and SD, by plotting TSS versus $(SD)^{-1}$. Again, the data are noisy, but consistent with a linear relation between the variates. Again, the physically-based zero-intercept LS line is as good a regression as the OLS, differing in standard error by only 0.5 %. The slope of this line is 19.6 based on 1206 data points, and is well within the range of literature values. For 400 data points from Galveston Bay, Ward and Armstrong (1992a) found a slope of 13.0, within but near the lower limit of the range of the literature values. We judge the Corpus Christi relation to be better based, and adopt this as a proxy.

During its three-year program, SWRI performed a number of measurements of fractional transmission for a 10-cm path with a transmissometer. From Beer's Law, with z=0.1 m, the extinction coefficient is given by

$$b = -10 \log(Tr/100)$$

Tr denoting percent transmission and log the natural logarithm. While turbidity T, whence SS, can be expected to be proportional to b, the proportionality constant is in doubt. Unfortunately, SWRI did not make any companion measurements of turbidity or suspended solids. TSS data collected by the TWC/TDWR SMN during the same time period in the open waters of Corpus Christi Bay (Segment 2481) suggest that a factor of 10 would be appropriate. The resulting equivalent TSS values along with those measured by the SMN are shown in Fig. 2-8.

Turbidity data determined by nephelometric methods also appear in the Corpus Christi data base, but unfortunately there are no paired measurements of TSS to determine the relation between the two. There are, however, paired measurements of both turbidity measures with Secchi depth. These data are displayed in Fig. 2-9. While there are certainly regression lines that can be fitted to these data, and the variable of SD^{-1} eliminated algebraically, the parameters of the lines prove to be sensitive to the order of regression (i.e., which variable is taken as independent), which introduces considerable uncertainty into the relation between JTU and NTU turbidity. It appears that a reasonable relation for Corpus Christi Bay is JTU = 1.86 NTU, but this should be regarded as provisional.

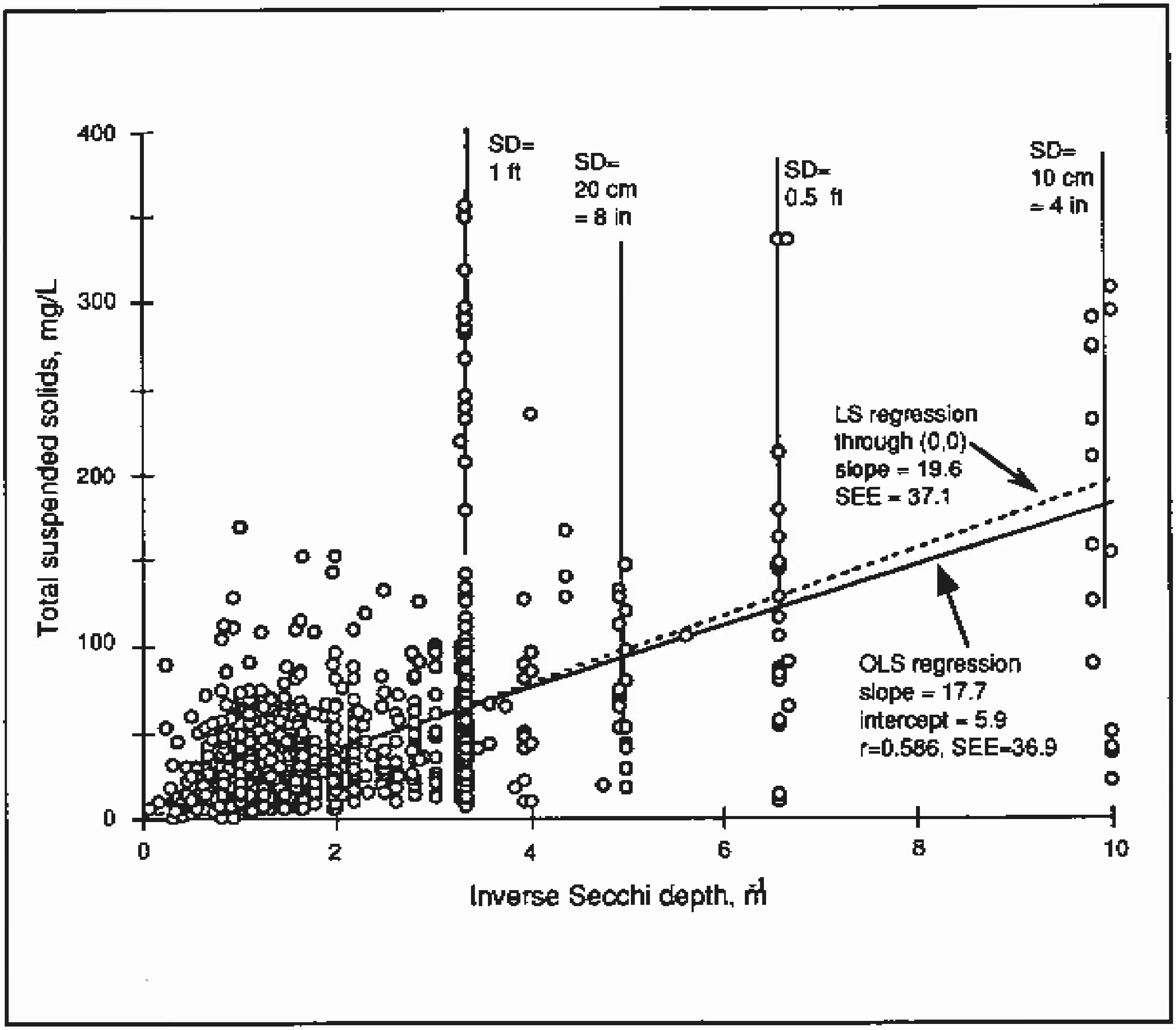


Figure 2-7. TSS versus inverse Secchi depth in Corpus Christi Bay

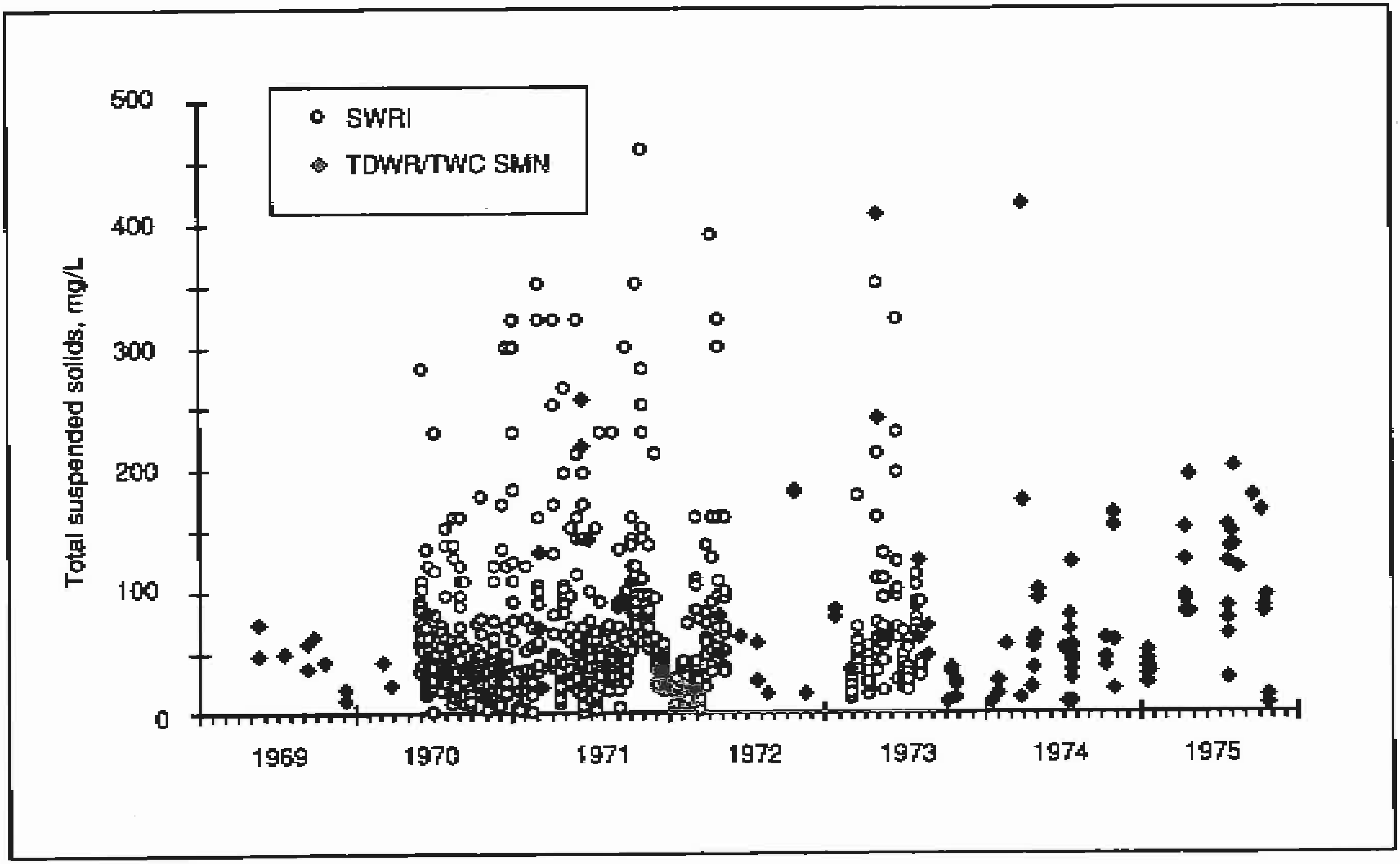


Figure 2-8. Comparison of SMN-measured TSS and value computed from SWRI data

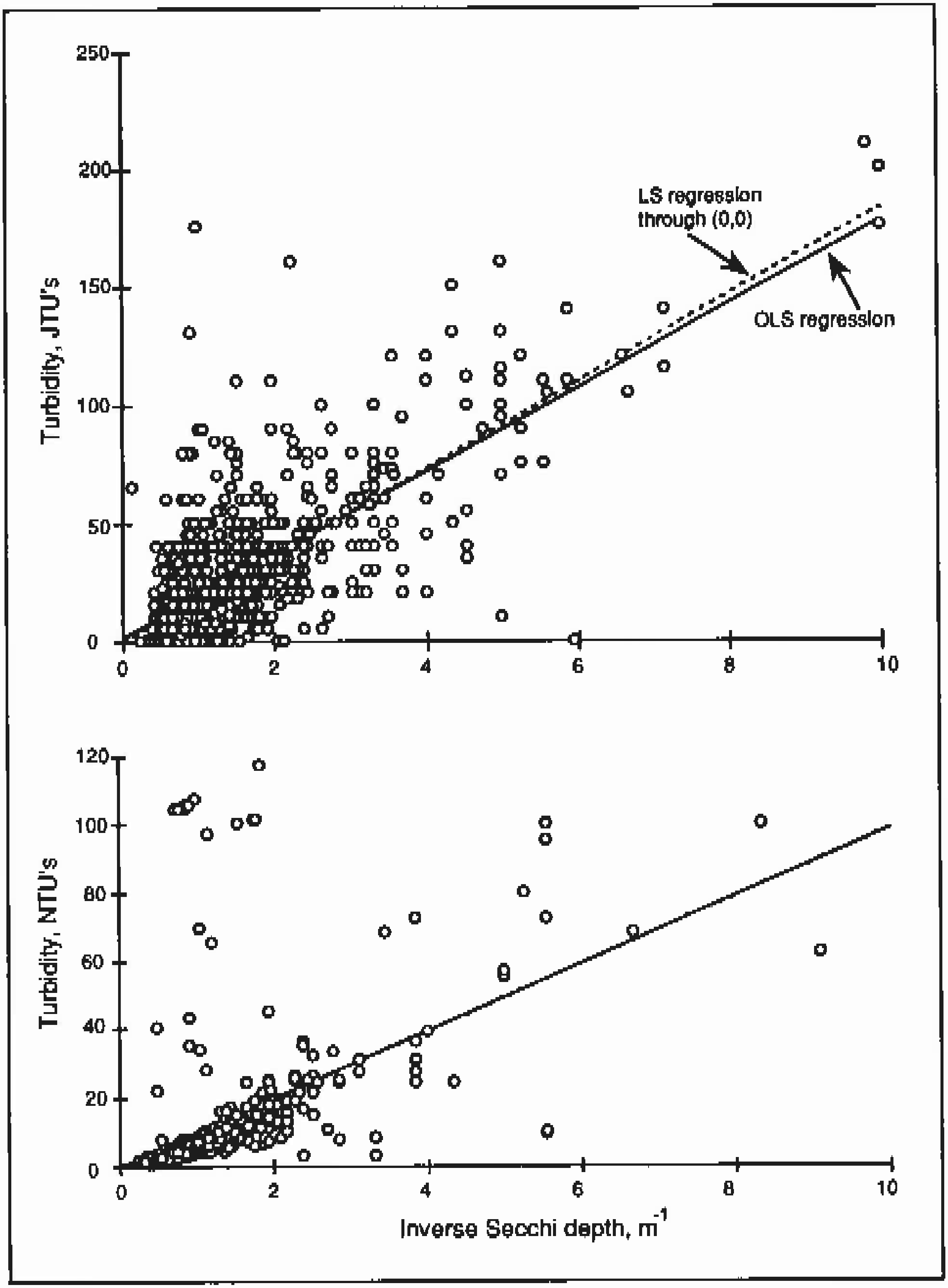


Figure 2-9. Turbidity in JTU and NTU versus inverse Secchi depth

In conclusion, for the present study, the following forms are adopted to serve as proxy relations giving TSS in terms of turbidity measurement, in order of priority (after, of course, the actual measurement of TSS):

$$SS = -10 \log\{Tr/100\} / z$$

$$SS = 0.93\ T \qquad (2\text{-}9)$$

$$SS = 1.73\ TN$$

$$SS = 20 / SD$$

where SS is suspended solids in mg/L, z is the optical path length in meters, Tr is percent transmission, T is turbidity in JTU, TN is turbidity in NTU or FTU, SD is Secchi depth in meters, and log denotes the natural logarithm. These relations are at best approximate and much is left unsettled.

2.4 BOD and related parameters

Biochemical oxygen demand (BOD), oil & grease, and volatile solids are tests which have developed from situations dominated by oxygen-demanding pollutants, and while their merit as water-pollutant parameters continues to be debated, the fact is that these parameters enjoy the longest period of record in most aquatic systems.

Since the classical work of Phelps and Streeter the biochemical oxygen demand (BOD) has become one of the fundamental parameters for estimating the presence of oxygen-demanding organics in a water sample (either from a sewage effluent or from a natural watercourse) and is one of the central parameters in the mathematical modeling of dissolved oxygen in the watercourse. Despite this long history of use, which can be traced back to the 19th Century, the BOD test is in many respects still controversial and is a continuous source of debate regarding the correct laboratory procedures and interpretation of the results. A detailed review is given in Ward and Armstrong (1992a), which is summarized here.

Fundamentally, the BOD is the amount of dissolved oxygen (DO) consumed in a sample of water during some period of time. The standardized test for the 5-day BOD is familiar and detailed in, e.g., *Standard Methods* (APHA, 1971, 1985) and HMSO (1983). We also note that the BOD, as established in a laboratory test such as *Standard Methods*, is a direct measure of a biochemical process at work within a stoppered bottle in equilibrium, a trivial observation at this point but which looms large in the interpretation of BOD data. The basic concept of BOD was to measure the potential oxygen depletion in a natural stream due to an injected waste. Since then the concept has evolved in two separate directions, both of which are referred to as BOD, to compound the confusion. The first is the oxygen consumed within the watercourse by the degradation of organic wasteloads, for which the ultimate BOD is the crucial quantity, with the oxygen depletion directly related through Phelps Law as described below. The second is the evolution of the

BOD bottle test as a measure of the organic wasteload of an effluent, and therefore, as a direct monitor of the operation of a waste-treatment facility and the key design parameter for treatment processes.

The amount of oxygen consumed as consequence of aerobic biochemical processes in a water parcel, whether it be a laboratory BOD bottle on a shelf or a moving parcel of water embedded within the flow of a natural watercourse, is directly dependent upon a number of variables, as follows:

(1) Types of bacteria present in the water;
(2) Initial quantities of each type of bacteria present;
(3) Multiplication or growth rates for each type of bacteria present;
(4) Chemical characteristics of the substrate, i.e., the oxidizable organic constituents within the water;
(5) The quantity, or concentration, of the oxidizable constituents;
(6) Constituents which act as an inhibitor or a stimulant for the bacterial metabolism;
(7) Environmental parameters, most notably pH and temperature;
(8) Other aerobic organisms in the water, notably phytoplankton.

It is apparent that there is a multiplicity of factors that can affect the BOD in the water parcel.

While there are several methods available for determining BOD, it appears (given the poor documentation) that almost all of the measurements from Corpus Christi Bay are based on the traditional dilution-series 5-day BOD either with natural seed or, for those performed by the Texas Department of Health Laboraory, with the cultured "Texas" seed.

The precision of a BOD test is exactly that of the dissolved oxygen measurement compounded by the number of independent DO measurements (there are at least two). The accuracy is another matter, particularly with respect to replicability. The manifold processes underway in a BOD bottle render the oxygen depletion quite complex and problematic. One ubiquitous source of error is in the dilution itself. In many studies employing BOD, the phenomenon has been encountered of increasing BOD (per unit volume) as the sample is subjected to greater dilutions. this is attributed variously to toxicity in the sample water, contamination of the dilution water, and selective stimulation of bacteria by the nutrients in the dilution water, among other hypotheses. In Espey et al. (1971), data from Galveston Bay on the ratio of BOD_5 in 4:1 dilution to that of an undiluted sample were tabulated, and shown to range from 1.5 to 4.5, with the highest values in Galveston Harbor and the lower bay. A possible explanation for the phenomenon lies in the Monod equation for bacterial growth (e.g. Monod, 1949), giving the growth rate as a function of substrate concentration C, which also governs the oxygen consumption rate. If C is initially high, the growth rate is essentially its maximum value. As C is reduced, the growth rate declines. The rate constant for BOD exertion is, therefore, a nonlinear function of the dilution factor, which of course contradicts the basic assumption underlying the dilution approach, that the BOD depletion is simply proportional to dilution. In the present data base, BOD is reported without any information on dilution, so we must regard the dilution factor as a (considerable) source of uncertainty in the measurement.

The controversy attending the BOD measurement has led some researchers to propose alternatives to the parameter. One such is the total organic carbon (TOC), a suggestion motivated by the notion that BOD measures the organic carbon substrate (Busch, 1966). However, the nature of the carbon compounds as well as the capabilities of the bacteria dictate the oxygen demand. For example, Maier and McConnell (1974) report TOC:BOD ratios in the range 0.5-1.0 for "simple" carbon compounds (e.g., glucose, stearic acid), and in the range 0.5-4.0 for wastewaters of varying treatment levels. Of course, if a relation between TOC and BOD could be established, it would serve as a proxy to extend the record for either parameter. There are some 435 paired measurements of BOD and TOC in the data base for the Corpus Christi Bay system, plotted in Fig. 2-10. It is readily apparent that there is little correlation: the explained variance (TOC on BOD) weighs in at 15.5%.

2.5 Nutrients and indicators of productivity

Nutrients have an ambiguous position in the assessment of water quality, in that they are necessary to support a healthy aquatic ecosystem, but in excess can lead to nuisance conditions such as hyperstimulation of primary production, in which case they are regarded as pollutants. Some of the earliest chemical measurements available for the Corpus Christi system address certain of these nutrients associated with waste discharges.

Two of the principal nutrients, nitrogen and phosphorus in their various forms, play an essential rôle in aquatic biological processes. Further, their concentrations can be significantly augmented by the activities of man, especially through point discharges of municipal and industrial wastes, and through runoff from modified watersheds, especially landscaped areas, agricultural operations using applied fertilizers, or ranching with concentrated herds. While nitrogen exists in four principal species, not all of these are routinely measured. The most frequently measured forms, and therefore the best data base, are ammonia and nitrate; though wherever possible we also present analyses of organic nitrogen and Kjeldahl nitrogen (the sum of ammonia and organic). Nitrite is much less frequently measured, and in the watercourse is rather unstable, being readily oxidized or reduced to other forms (and therefore nitrite concentrations are much less than those of the other species). As the relative proportions of these species are a strong function of origin, transport and microbial kinetics, no relation among them is meaningful for serving as the basis of a proxy equation. (The relations typically used in mass budgeting, e.g. Meybeck, 1982, are strictly applicable to natural waters, not waters potentially subject to abnormal sources of nitrogen or abnormally vigorous populations of chemautotrophs.) Therefore, separate analyses were carried out for each of these. It should be noted that the measurement of organic nitrogen refers in fact to the least refractory components of the organic pool (McCarty et al., 1970).

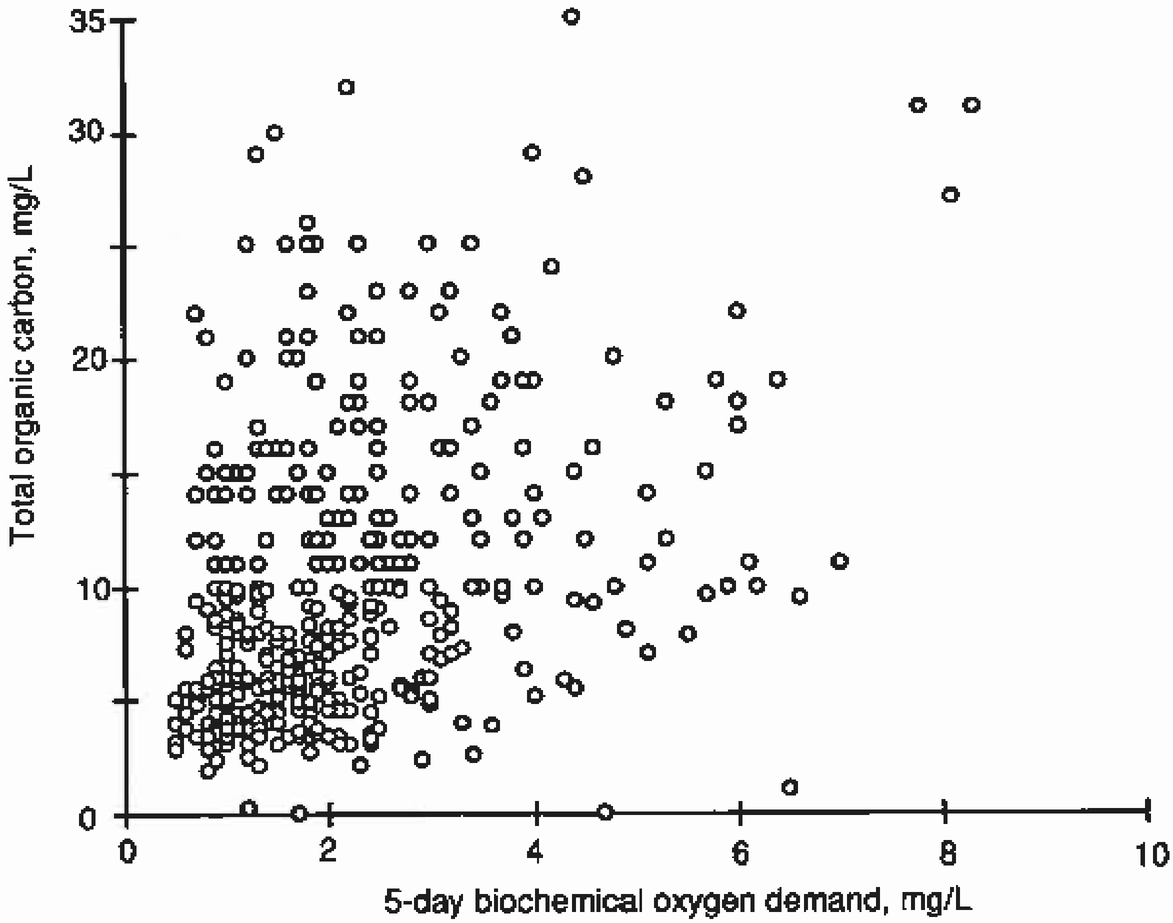


Figure 2-10. TOC versus BOD, paired measurements in Corpus Christi Bay system

The most common measures of phosphorus concentration in the Corpus Christi data base are orthophosphates and total phosphorus. As is the case with nitrogen species, there is no consistent relation among the forms of phosphorus that would serve as a proxy. The proportion of orthophosphate may be quite low in natural waters, but as much as 90% of the total phosphorus in municipal effluents. Also, the natural concentrations in seawater are much different from those in fresh water. Accordingly, we have chosen to focus the analysis specifically upon total phosphorus, although statistical results are presented for all of the phosphorus forms represented in the data base. One potential source of uncertainty in this measurement is the treatment of particulate (versus dissolved) phosphorus. Phosphorus is sorptive and has an affinity for fine-grained suspended sediments. In some of the data sets, it is not clear whether the total-phosphorus analyses are restricted to the dissolved fraction (i.e., whether the sample is filtered) or includes the particulate. Moreover, the sorptivity diminishes with salinity, so the distinction between filtered and unfiltered analyses becomes less important at higher salinities. Dr. Terry Whitledge (pers. comm., 1996) notes that he has compared filtered and unfiltered analyses of these nutrients and found no significant differences, in which case the lack of information on filtering would be immaterial. (Dr. Whitledge observes that this is yet another instance of freshwater dogma being imported into estuarine science.)

As a surrogate indicator of phytoplankton biomass, chlorophyll-*a* has become a common water quality parameter. There are two basic analytical methods applied to water sampling in the Corpus Christi Bay data, spectrophotometric and fluorometric. The TNRCC SMN data are indicated to be the former. (An older method, the trichromatic, was employed in the 1960's and 1970's by the State, but this STORET code does not appear in the Corpus Christi data.) A related parameter is the metabolic product phaeophytin, sometimes loosely referred to as "dead" chlorophyll. This has been measured by the TNRCC SMN program only. Table 2-2 summarizes the statistical association between the two in the Corpus Christi data (for those pairs for which both parameters were above detection limits). It is noted that the coefficient of variation of both chlorophyll and phaeophytin exceeds 200%, no doubt due to the significant space-time variability of phytoplankton in Corpus Christi Bay as well as imprecision in the measurement, and that chlorophyll-a and phaeophytin are virtually uncorrelated. (This is not surprising, since one does not expect *a priori* a correlation between the two.) One other chlorophyll measure in the Corpus Christi data base is that of chlorophyll-b, reported by the SWRI monitoring program of the early 1970's. This program did not measure chlorophyll-a. Chlorophyll-b is a much less useful pigment than chlorophyll-a, it occurs in much smaller concentrations than chlorophyll-a, and methods of determination are less reliable (APHA, 1985, 1989). Why only chlorophyll-b would have been measured is inexplicable (see Chapter 5). This data was not used in the present analyses.

Silicon in its various forms is an important nutrient, especially in supporting the silaceous cells of diatoms, an extremely significant component of the phytoplankton in estuaries. Silicon is derived primarily from weathering of terrestrial landforms and is brought into the estuary by runoff and inflow. In some estuaries, silicon can be as useful a watermass tracer as salinity, see Ward

TABLE 2-2

Statistics of chlorophyll-a and phaeophytin data in SMN data base for CCBNEP study area

Number of paired observations		2312
Number of paired observations above detection limits		730
Chlorophyll-a		
	mean	21.96
	standard deviation	54.4
Phaeophytin		
	mean	6.76
	standard deviation	15.1
Linear correlation: phaeophytin on chlorophyll-a		0.289
Explained variance		8.3%

and Montague (1996) and references therein. It has been used for this purpose in a study of the circulation of Nueces Bay by Whitledge (1993).

Total organic carbon is an ambiguous parameter, measuring a portion of the carbon nutrient pool, organic pollutants (cf. the discussion of BOD in the preceding section), and the end product of primary production. Its interpretation as nutrient, pollutant, or base of the food chain therefore must be tempered by the concentrations of other parameters.

2.6 Trace elements and organic pollutants

Trace elements, primarily heavy metals, and volatile organics, including pesticides, are more recent arrivals on the analytical scene. The EPA Priority Pollutant List, per Section 307 of the Clean Water Act, is a means to specify which individual compounds are to be given detailed study due to their high pollutant or toxicological potential, a list which continues to mutate as constituents are listed and de-listed by EPA (see Keith and Telliard, 1979, Finnigan, Hoyt and Smith, 1979). The procedures for measuring these constituents are collectively referred to as "instrumental" methods, in contrast to, say, traditional stoichiometric methods which introduce reactants in known quantities to infer the presence and quantity of the analyte. All instrumental methods require calibration of the instrument to prepared, known standards, in the process for which the reliability of the method as a function of standard concentration is quantified. (This is discussed further in the context of analytical uncertainty in Chapter 4.)

By far, the most common method for elemental analysis used for the measurements in the CCBNEP data base is spectrochemical, consisting of the following generic steps:

- physical alteration of the sample to facilitate radiative behavior
- electromagnetic radiation exposure of or emission by the sample
- monitoring of electromagnetic radiation as a function of frequency
- recording and interpretation of the line spectrum

Atomic absorption methods monitor the selective absorption of EM radiative energy, generally in the UV and visible bands, by the sample. Atomic emission methods monitor the EM radiation emitted by the sample after non-radiational excitation, either thermal or electrical. Atomic fluorescence methods monitor EM radiation emission after excitation by radiation. The principal difference between different AAS methods derives from how the sample is atomized. The simplest is probably flame AAS (FAAS), in which the sample is nebulized into a flame, thus dissociating compounds into atoms. A more accurate procedure is graphite furnace AAS (GFAAS) in which the sample is heated to atomization in a small graphite tube, or "furnace." In either method, the atomized sample is exposed to radiation from a specially designed lamp.

AES methods also mainly differ in the treatment of the sample, i.e. what the source of energy is that "excites" the sample atoms to emit radiation. Generally, this energy source accomplishes both the atomization of the sample and its excitation, in contrast to AAS methods for which both an atomization energy source and and radiation source are needed. Older AES techniques employed flames (thermal sources) and sparks or arcs (electrical sources), but newer, more accurate procedures use plasma sources. Because of the similarity of the FAES procedure to the FAAS, many instruments can be operated in either configuration. A similar recent hybrid of AES using a graphite furnace source is referred to as GFAES. The most important plasma methods are the inductively coupled plasma (ICPAES), direct current plasma (DCPAES) and microwave-induced plasma (MWIAES). Details of all of these methods are summarized by Vandercasteele and Block (1993).

For present purposes, the two essential facts of these types of spectrometric analyses are: (1) choice of method governs the range of elements that can be detected and the sensitivity of the measurement; (2) no information is preserved as to the compounds in which the element occurs in the sample. With respect to the former, there is generally a trade-off between sensitivity and range (and with cost). With respect to the latter, the "speciation", i.e. the specific compound in which the element appears, can be extremely important in determining the nature of the chemical and biological effects of the element. While there is a general philosophy that larger concentrations of different species are in some way associated with larger elemental concentrations, lack of information on speciation is often a major impediment in interpreting metals data.

In order to analyze the concentration of an organic compound, first a separation of the mixture of compounds in the sample must be achieved. The predominant technique for this is to pass the sample through a chromatograph column, a glass tube coated with thin liquid film, or packed with liquid-coated granular solids. Passage of the sample through the column is facilitated by a carrier gas. The time required ("retention time") for a compound to traverse the column ("elution") depends upon the partitioning of the compound between the carrier and stationary medium, and is generally unique to the compound. The pattern of signal (proportional to concentration) versus retention time (the "chromatogram") is determined by injecting standards of known composition.

A more versatile operation is to couple the output from the chromatograph column into the ion source of a mass spectrometer, referred to as GC-MS. The mass spectrometer determines the rate of ion flow versus the mass of the ion from which specific ions can be identified, hence the, perhaps inappropriate, name "mass spectrum." (See Chau and Afghan, 1982, Cairns and Sherma, 1992, and Vandecasteele and Block, 1993, for details of MS operation, and *Standard methods*, e.g. APHA, 1985, 1989, for set-up procedures.) Quantification again requires a calibration standard, and identifying mass spectra requires access to reference spectra. (A mass spectrometer as a method of ion identification needs only to be coupled to an ion source. The same procedures used to atomize samples in AAS and AES can be coupled with an MS for elemental analyses.) Both chromatograms and mass spectra are produced by GC-MS. The high data generation rate and the need to access reference libraries of known mass spectra have led to coupling the GS-MS to computers, which, in turn, has led to a considerable automation of the GC-MS procedure, including software for searching and identification as well as quantifying the resultant compound concentrations. Such automated analyses can produce erroneous results if not carefully monitored and cross-checked by competent chemists (see, e.g., Swallow et al., 1988, Alford-Stevens et al., 1988).

The essential facts of GC-MS and related procedures in the context of interpreting data from Corpus Christi Bay are: (1) the sensitivity and accuracy of the method, even when highly automated, are determined by the skill of the operator and the care used in set-up and sample preparation (including how the column is packed); (2) generally, the GC-MS output is searched for specific compounds, according to specified analytes and available information on the chromatogram and mass spectra of those analytes, that is, the method does not produce a census of all compounds in the sample; (3) compounds can be missed, inaccurately identified, or erroneously quantified if they are not adequately separated in elution from the column or if their mass spectra are insufficiently differentiated.

Any analytical method has a limit to its resolution, i.e. to its capability to detect small differences in concentration. The ability to quantify elements or compounds in trace concentrations is particularly limited by the resolution of the method. Part of the calibration and standardization procedure of the above methods is empirical determination of the resolution, expressed as a "detection limit." This is discussed further in Chapters 4 and 5 in the context of data processing and interpretation. For now, it should be noted that the AAS, AES, and GC-MS techniques in fact produce a *number* for a desired element or compound concentration. If this number is considered smaller than the detection limit of the method, the number is deemed to be fictitious, and is replaced by the statement that the substance is "nondetectable", meaning at a concentration below the empirical detection limit. It has been argued that information is sacrificed in "censoring" the data in this fashion (see, e.g., Porter et al., 1988, D'Elia et al., 1989), but the convention is ingrained in current analytical procedures.

For the CCBNEP data base, 48 organic compounds and 31 metals for the water phase were identified for specific compilation and analysis, as listed in Table 2-1. Data for additional compounds, such as the entire suite of priority pollutants and other metals, are in fact available for the system, but the number of such measurements are so few, and the number of concentrations above detection limits fewer yet (almost always zero), that these compounds and elements were excluded from further consideration. Even at this, the data base for many of the compounds listed in Table 2-1, as will be seen, are too sparse to support meaningful analysis. Generally, the metals are analyzed by one of the elemental instrumental methods, usually AAS, while the organic compounds are analyzed by GC-MS and related methods. The organics may be broadly categorized as petroleum-derived hydrocarbons and pesticides.

The utility of these parameters for water-quality analysis continues to be vexed by complex and uncertain analytical procedures. Especially for trace organics (a.k.a. micropollutants, organo-xenobiotics), protocols and procedures are still evolving, and this is reflected in a confusion of data acquisition. Some of the problem originates in the multiple forms a specific organic can assume: various isomers, analogs and metabolites. Further, the nomenclature for many of these is nonstandard and contributes to the confusion, particularly in data reporting. Lindane, for example, is reported variously as gamma-BHC (benzene hexachloride), gamma-HCH (hexachlorocyclohexane), as well as commercial names such as Lindafor, Lindagram and Lindamul (EPA, 1988). (Most of these problems were mooted in this project because the amount of data available was so limited that little meaningful analysis could be performed.)

The term "total" acquires a new dimension when applied to trace organics, in that it may mean the total of all isomers and analogs for a given compound, or the total of the compound and its metabolic products (with or without all isomers), or the total of an entire class of organic compounds. There is no universal convention, and the meaning of "total" must be considered specific to the intentions of the agency generating the data.

The principal concern with organochlorine pesticides, which include lindane, methoxychlor, endrin, endosulfan, dieldrin and heptachlor, is their long persistence in the environment. In contrast, the organophosphates like parathion and malathion are probably more toxic but also much more short-lived. The insecticide dichlorodiphenyltrichloroethane (DDT) is certainly the most prominent of the organochlorines (a.k.a. chlorohydrocarbons) and the one for which the available data base in Corpus Christi Bay is greatest. The use of DDT was banned in the U.S. in 1972. There are several metabolic products, the most important of which (and the only two considered in this compilation) are DDE and DDD (a.k.a. TDE); others include DDNS and DDOH. DDT as a technical product is comprised of as many as 14 analogs and isomers. By far the most important are p,p'-DDT and o,p'-DDT. The relative proportion of the two is a function of the proportion in the initial source, and of the relative kinetics and metabolism in the receiving water. Neither of these proportions is particularly well-defined, though the former is probably better established than the latter, to be about 70% p,p'-DDT and 20% o,p'-DDT in technical grade DDT (Buechel, 1983). This is roughly consistent with the rule-of-thumb of a 3:1 ratio of p,p'-DDT to o,p'-DDT that seems to be current now. Both forms are hydrophobic and sorb readily to fine particulates, both sediments and phytoplankton (Crompton, 1985). Apparently microbial assimilation is stimulated by sorption (Chau and Afgan, 1982, Crompton, 1985) but appears to affect both isomers equally. Treatments of the kinetics (including volatilization) of DDT make no differentiation between the isomers (e.g., Moore and Ramamoorthy, 1984) so we assume that their ratios will be preserved in the receiving water and sediment. Many agencies report simply total DDT. Usually this means the total for all isomers. On the basis of the above proportion, the relation between total DDT and p,p'-DDT is taken to be:

$$\text{Total DDT} = 1.4 \cdot (\text{p,p'-DDT}) \qquad (2\text{-}10)$$

While this appears to be a workable proxy relation, we have no reported paired measurements by which we can test it. In the subsequent analyses, we present separate treatments of the total DDT data as reported, the individual analog analytes, and the extended "total DDT" data set (designated WQ-XDDT and SED-XDDT) using the above proxy relation.

In the CCBNEP data base, the pesticides best represented after DDT are toxaphene and chlordane. Each is an insecticide, and each is a mixture of compounds whose analysis is confused by poorly defined composition. Toxaphene is a mixture of chlorinated camphenes, estimated to be made up of at least 180 separate compounds (Rice et al. 1986, Cairns and Sherma, 1992). The fact that its principal compounds degrade ("weather") at different rates complicates the analysis and identification problem, It was the insecticide of choice after the ban on DDT, until it was also banned in 1982. Technical chlordane, an organochlorine, includes two isomers, cis- and trans-chlordane (gamma-chlordane), as well as heptachlor and heptachlor epoxide, each of which is a pesticide in its own right. Chlordane is toxic and persistent in the environment, with an estimated half-life of 5-15 years. The principal degradation product is nonachlor (several isomers, Chau and Afghan, 1982). Chlordane is metabolized in most organisms to two epoxides (heptachlor epoxide and oxychlordane) but the degree of metabolization is highly variable (Dearth and Hites, 1991). Since the mid-1980's, the use of chlordane as well as the individual component pesticides has been widely curtailed (EPA, 1988) and the use of chlordane was finally stopped in 1988.

Polychlorinated biphenyls (PCB's) is a class of biphenyl compounds containing up to 10 chlorine atoms (10 being the maximum number of sites available). There are 209 possibilities ("congeners"), and the horrible formulaic names are avoided by international convention of numbering the congeners from 1 to 209 (Alford-Stevnes, 1986) increasing with the level of chlorination. In the United States, virtually all PCB's were manufactured by Monsanto, under the trademark name Aroclor. PCB's are not pesticides *per se*, but are usually addressed with pesticides because of their similar chemical nature and analytical signatures (e.g., Murty, 1986). (In fact, PCB's can interfere with the determination of pesticides, and it is possible that early pesticide data, prior to about 1970, included poorly differentiated GC peaks due to PCB's and are therefore overestimated.) PCB's in fact had a wide range of commercial uses, from inks and NCR paper to fire retardants and dielectrics (Hutzinger et al., 1974, Erickson, 1986). Because of this, PCB use became widespread, until 1976 when Congress banned their manufacturing and sale, and limited their use to totally enclosed systems such as transformers. They are still widely used in capacitors, transformers and electromagnets.

PCB's are frequently analyzed by component Aroclors, which are identified by 4-digit product numbers, e.g. Aroclor 1221. In fact, PCB's often referred to interchangeably as Aroclors, a convention that is fortunately declining, because the two are not equivalent. Each Aroclor is a complex mixture of PCB's and other compounds (Hutzinger et al., 1974). (There are even a few Aroclors that are non-PCB's.) The Aroclor compounds are identified by the *pattern* of peaks as the PCB's in the mixture elute from the GC column. But this chromatogram pattern is complex and can be obscured by GC techniques with poor resolution. Even when supplemented by a sensitive detector, the analyst must judge the identity of the compound by its component retention times, in comparison to a standard chromatogram, and this judgment can be confused by other compounds with similar retention times. (Pattern-recognition software is now available that performs this judgment; opinions vary about its ability to reduce the margin of error, Schwartz et al., 1987, Swallow et al., 1988, Alford-Stevens et al., 1988.) The identification accuracy is improved with a mass spectrum. When a mixture of Aroclors is present the problem becomes even more complex. To make matters worse, the components of an Aroclor mixture degrade in the environment at different rates, so the resulting chromatogram may not agree with the standard, and therefore be liable for misidentification (Schwartz et al., 1987, Alford-Stevens, 1986).

It is much preferable to analyze and report PCB's as individual congeners. Since there are potentially 209 such congeners, this can be a major analytical effort. It is made more demanding by the fact that different methodologies may be needed to resolve the entire range of retention times. The analytical effort can be moderated by limiting the analysis to a specific subset of the congeners. For example, Schwartz et al. (1987) analyzed 105 isomers, Baker and Eisenreich (1990) analyzed 35, Bergen et al. (1993) analyzed 13. Another means of reducing the analytical effort is to analyze and report PCB's by levels of chlorination, i.e. by isomer groups. In this approach, a representative congener for each chlorination level is used to calibrate the MS for all isomers in that chlorination class; ideally, the representative is near the mean detector response for its class.

PCB's are ubiquitous pollutants. This is due to their widespread use, but also to their mobility. Like chlorinated pesticides and many other organics, they are hydrophobic, have an affinity for very fine-grained particulates, and are leached from landfills, carried by runoff from watersheds, and transported by wind. They are persistent, their half-life generally increasing with level of chlorination, and many are toxic. The atmosphere is a particularly important means for PCB transport. (A recent review of the PCB literature, emphasizing analytical procedures, is provided by Erickson, 1986.)

Another important class of trace organics is polynuclear aromatic hydrocarbons (PAH's), which include polycyclic aromatic hydrocarbons as a subset. (Some authorities refer to the polynuclear class as PNA's and reserve PAH specifically for the polycyclic compounds, e.g. Bjørseth, 1979.) Most studies focus on a relatively small number of "indicator" compounds, primarily to simplify the analytical effort. These usually are a subset of the following:

fluoranthene	pyrene
anthracene	benz(a)anthracene
benzo(b)fluoranthene	benzo(a)pyrene
benzo(e)pyrene	acenapthene
perylene	benzo(ghi)perylene
fluorene	phenanthrene
chrysene	coronene

Probably the most important physicochemical aspect of a PAH is its molecular weight. The light end of the range includes napthalene ($C_{10}H_8$), fluorene and anthracene, and the heaviest, benzo(a)pyrene, chrysene and coronene ($C_{24}H_{12}$). Sometimes PAH's are analyzed and reported in molecular weight classes.

PAH's are introduced into the environment as combustion products, especially from fossil-fired power generation, emissions from petrochemical operations especially petroleum catalytic cracking and the production of asphalt and coke, and through release and degradation of petroleum compounds. Like PCB's, PAH's are widely dispersed in the atmosphere, and enter aquatic systems through rainout and fallout. They are directly injected through waste discharges and spills. (There are also natural sources of PAH's.) The relative distribution of concentrations, or the ratios of pairs of PAH's, have been used as a "pattern" to identify the source of the PAH's (Bjørseth, 1979, Neff, 1979). This obviously is made more difficult when many multiple sources are involved. Moreover, kinetic processes in the environment can modify the component PAH's and obscure their original pattern.

PAH's vary in their kinetic behavior in the environment. Some, like pyrene, perylene, anthracene, and naphthacene, are readily photolysed, but others like phenanthrene and chrysene are unaffected by exposure to light. Some, especially the higher molecular weight species, are readily oxidized, and react with oxides of nitrogen and sulfur. PAH's are hydrophobic and sorb to particulates. There is some indication that when sorbed, PAH's are more stable and less likely to be kinetically degraded. A useful, comprehensive review of PAH's in aquatic environments is presented by Neff (1979).

The environmental concern with PAH's has focused mainly on their features as carcinogens, especially the higher molecular weight species. One of the most carcinogenic is benzo(a)pyrene (BaP), which has received considerable study in the past, and is sometimes used as an indicator for the entire PAH class. Although PAH's are considered to be toxic, especially the lower molecular weight species, there is relatively little information extant on their toxicity to marine organisms.

Frequently, agencies and analysts report "total" PCB's and "total" PAH's. This means the arithmetic concentration sum of the compounds analyzed. If six Aroclors are sought, the total PCB's will be the sum of concentrations of these six. If eight chlorination levels are determined (out of ten), total PCB's means their sum. If 35 congeners are analyzed, total PCB's means the sum of the 35 concentrations. Clearly, "total" is a relative term, and cannot be presumed to be comparable between two different sources of data. On the other hand, if we have reason to believe that the representation of PCB's is dominated by a few Aroclors, then it may be appropriate to combine data on total PCB's analyzed by these Aroclors with totals based upon, say, chlorination levels. Caution must be applied, however, lest apples be mixed with oranges. Similarly, with PAH's the "total PAH" is the arithmetic sum of the compounds analyzed. Neff (1979) observes that the lower molecular weight PAH's, which are more volatile and soluble, are not handled efficiently by many collection and extraction techniques, and therefore the "total PAH" reported by many labs do not represent the low molcular weight species. Some of the data sources for Corpus Christi Bay analyze only 4-6 PAH's, while others determine an extensive suite. One data source, the EMAP/REMAP program of EPA, presented two different "total PAH" values, corresponding to different suites of measurements (dictated by the practice of two laboratories); these totals often differed substantially. Parameters for total PAH and total PCB are included in the CCBNEP list. However, data sources for these were limited to those that reasonably approximated the same analytes.

2.7 Coliforms

The reader has no doubt noticed that this chapter has addressed water and sediment parameters in general order of increasing uncertainty. Therefore, it should be no surprise that coliforms are treated last. The specification of two basic classes of bacterial growth-response referred to as "total coliforms" and "fecal coliforms" is a controversial, low-precision measure, originally intended to provide an index to the extent of contamination by pathogens of enteric origin. There is, due to the extensive aquatic recreational activities as well as shellfish harvesting, a considerable data set for Corpus Christi Bay.

Sometimes, total coliforms are measured, sometimes fecal, occasionally both. The question is whether there is a stable relationship between the two that will allow us to proxy a data set for one or the other. To the extent that both are dominated by an origin in discharge of sewage, the answer would be anticipated to be affirmative. However, both—especially total coliforms—are the result of a large, varied community of microorganisms with various non-sewage, non-anthropogenic, and even non-mammalian sources. (This, in fact, is the nub of the controversy surrounding the efficacy of coliforms as an indicator organism.) This question was explored by Ward and Armstrong (1992a) using data from Galveston Bay for which an extensive data base exists for paired measurments of both total and fecal coliform.

There is a rule-of-thumb about, that fecal coliforms are approximately one-fifth of total coliforms (e.g. Kenner, 1978) but there seems to be little published support. Ward and Armstrong (1992a) argue that this relation arises from the uncritical use of statistics. In the Galveston Bay data, Ward and Armstrong (1992a) present a regression analysis, in which

$$F = 0.352\,C + 1056$$

(where F denotes fecal coliforms and C total coliforms per 100 mL) with a correlation coefficient of $r = 0.889$. However, by the nature of the coliform test, $0 \le F \le C$. This relation induces a spurious correlation. If F and C are completely random variables, with uniform probability distribution, bound only by this inequality, Ward and Armstrong (1992a) show there is an "artificial" linear correlation of 0.65. For real coliform data, it is even worse, since the natural distribution is highly skewed, with most of the data clustering in the smaller values. Ward and Armstrong (1992a) assumed a lognormal distribution and computed $r^2 = 0.75$. With this much spurious correlation, the $r^2 = 0.79$ of the Galveston Bay coliform data is seen to have no significance at all. Instead, they examined the ratio F/C, and found that this ratio has a mean of 0.325 and a standard deviation of 0.306. For practical purposes, this ratio approximates a uniform distribution over the range 0 to 1. That is, the ratio is totally random. They conclude that there is no useful ratio by which fecal and total coliform may be related, and the two should be treated as independent measures. This same conclusion should apply as well to Corpus Christi Bay.

3. SEGMENTATION

As an estuary, the Corpus Christi Bay system is a watercourse that is transitional between a freshwater and a marine system. This transitional character implies a substantial gradient in the bay environment, as reflected in the water quality and sediment quality parameters addressed here. In order to carry out a trends analysis on the bay, the aggregation of an enormous amount of data is necessary, but in order to exhibit the spatial variation in these data, they must be disaggregated in some way to be representative of geographical position. One time-honored approach is to subdivide the system into segments. The specific strategies of water-quality segmentation are considered here, preliminary to the formulation of criteria and delineation of segments in Corpus Christi Bay. Reference is made to Ward and Armstrong (1992a) for a more extensive discussion of segmentation in estuary water-quality management.

3.1 Purposes of segmentation

Segmentation refers to the subdivision of an estuary into regions, and represents a compromise between the resolution of physical detail in the natural system, and the expediency of dealing with a small number of geographical units. Any segmentation system therefore entails a coarse level of spatial aggregation. The question is how coarse a resolution can the objective of the analysis tolerate, and therefore how small can the number of defined segments be.

There are two broad objectives for imposing a segmentation system on an estuary: administrative and analytical. The administrative objective refers to administration of laws and regulations. Therefore, part of the criteria for an administrative segmentation is an alignment of segment boundaries with jurisdictional boundaries, which can include:

State boundaries
county and district boundaries
state tract boundaries
geographical boundaries

For a large watercourse like Corpus Christi Bay, the segment boundaries can also reflect boundaries that are readily identifiable in the field, such as narrows, passes, and bridge crossings, and can also be based on the need for efficient access to the region for inspection or enforcement purposes. Therefore proximity to marinas and boat docks, or to highways and bridge crossings can form part of the criteria for segmentation.

The second broad objective of segmentation, *viz.* analytical, refers to the aggregation and analysis of data of some sort from the subregions of the bay. This segmentation is related to the nature of the data (or, equivalently, the objective of the analysis). Economic or demographic analyses will require different spatial aggregation, hence different segmentations, than, say, geological or climatological analyses. Independent of the nature of the data, it must be emphasized that the imposition of a system of segmentation is a compromise between some minimum level of spatial resolution (which carries with it a statistical level of confidence) and a minimum number of spatial units for analysis.

The definition of segments for an estuary becomes especially complex when the property of interest, say distribution of an organism in the bay, is dependent upon another variable, say water quality, which has its own spatial variability. Further, a segmentation system may in fact reflect both broad objectives listed above. The regulation of shellfish harvesting and the regulation of water quality, for example, have both an administrative objective, which may subject the segmentation to criteria of political boundaries and field-operations efficiency, and an analytical objective, which may require delineation of spatial variability of the target parameters.

One of the earliest, and therefore best-known, approaches to segmentation of an estuary for water quality purposes is that of Bostick Ketchum (1951a,b), who subdivided the estuary into segments of length equal to the tidal excursion. His segmentation is hydrographic in principle (see Section 3.2), and is based upon two fundamental postulates: (i) advection by the tidal current is the dominating transport, (ii) mixing is complete over each segment during each tidal cycle. On closer consideration, it will be seen that these two postulates conflict, in that to the extent that one is satisfied the other is violated. Of course, Ketchum's segmentation was devised to support *computational* analysis, which frequently imposes some rather strong conditions on the segmentation.

The most prominent example of segmentation for computational purposes is the gridding of a numerical model. Such segmentations basically observe the same philosophy stated above, in that a computational segmentation is a compromise between the need for a fine resolution of physical and water quality detail, and the need for as few a number of segments as possible in order to minimize computational overhead. However, the computational scheme imposes conditions of its own. For instance, the actual location of physical boundaries is altered to conform to the position of computational elements. For a finite-difference grid, such as that developed for Corpus Christi Bay by the Texas Water Development Board (TDWR, 1981), the segments are square regions one-nautical-mile on a side. A finite-element model repairs this geographical distortion to some extent (e.g., Klein and Ward, 1991, Longley, 1994), but still replaces the shoreline with straight-line segments. Further, in either type of numerical grid, the concentration of constituents is taken to be homogeneous within segments. (Actually, the mathematics of mass budgeting may assume some spatial distribution across a segment, linear in low-order finite-difference models, and linear or parabolic in finite-element models, but when the model is applied to real data, e.g. in validation, the data are generally averaged across the segment.) In general, however, a computational segmentation observes different criteria and is not as effective a schema for the analysis and depiction of water quality data as a segmentation expressly formulated for this purpose.

The General Land Office, and several other state agencies, employ the state tract system for segmentation. This is an example of a segmentation system that is purely administrative, and in which the constraints of operational surveying completely determine segment boundaries. The Texas State Department of Health employs a rather gross segmentation of the bays for monitoring and regulating shellfish harvesting. The segments generally correspond to large geographical subdivisions of the bay and have little correspondence to hydrographic or water quality features of the system. Again, it is a system devised for its administrative benefits, rather than analysis of water quality.

The most important administrative water-quality segmentation system is, of course, that of the Texas Natural Resource Conservation Commission (nee Texas Water Commission). The Corpus Christi Bay system, including the tributaries, is subdivided into nineteen segments. The TNRCC WQ Segments (also referred to as Classified Segments or Designated Segments) represent one of those instances of a segmentation system that reflects both objectives named above, i.e. it is used both for regulation and for analysis. In the regulation arena, the Water Quality Segments are the basis for setting water quality standards, hence underlie discharge permitting, compliance enforcement, and administrative actions. In the analytical arena, the Water Quality Segments are the basis for establishing monitoring stations and determining ambient water quality. The rationale for TNRCCC WQ segmentation boundaries is a combination of geography, tradition and politics.

The requirements of the CCBNEP Work Statement is that status-and-trends analyses be carried out for each of the Texas Natural Resource Conservation Commission Water Quality Segments in the Corpus Christi Bay system. However, to secure the objectives of this project, it is necessary to perform analyses on a finer spatial scale than possible with the TNRCC segments. Therefore, we have devised a system of "Hydrographic Segmentation" for Corpus Christi Bay to form the basis for detailed analysis. The criteria underlying the formulation of this (or, in general, any) analytical segmentation are developed by Ward and Armstrong (1992a) and summarized below, prior to presentation of the segmentation schema itself.

3.2 Principles of water quality segmentation

Just as different data collection programs have different objectives which inform the procedures and methodologies, so also are the sampling areas and sampling stations in general different from one agency to the next. Yet in many areas of the bay, *to within a certain level of confidence* (in the statistical sense), there is no difference between measurements taken at one position and those from another, perhaps even several kilometers removed. From the standpoint of identifying temporal trends in water quality and in characterizing regional water quality within the system, it is desirable to aggregate sampling stations from different programs. (Indeed, even within the same program, the same sampling station is not occupied precisely from one sampling run to the next.) With different data sets so aggregated, a sufficiently extended and dense set of data may be created to allow statistical characterization of these specific water quality regions.

The basic principle of such aggregation can be stated succinctly as follows. Aggregation of data should be based upon the determination of regions of homogeneity (within some statistical threshold), and zones or loci of sharp gradients in properties. The former should correspond to the *interior* regions of segments and the latter to *boundaries* between segments.

In order to minimize errors introduced by spatial aggregation, and to maximize its physical significance, the areas in which sampling stations are to be aggregated must be carefully delineated. This delineation should take into account transports, bathymetry, waste sources (where appropriate), inflows, and in general the distribution of physicochemical features which will either homogenize the parameter (to define the region encompassed by a water quality segment) or create steep gradients (to define the boundary between segments). It is useful to formalize these notions as specific criteria of segmentation, both to guide the specification of segments for the bay, and as a means of evaluating the suitability of existing agency segmentation systems.

Since water quality is a property of the fluid medium, one of the determinants of water quality is the pattern of transport within the estuary system. Therefore, variables which must be included in the definition of water quality segments are morphology and hydrography, *viz.*:

(1) Morphology: constraints on, or barriers to, flow and exchange:

- (1.1) Physiography should comprise the principal boundaries of segments wherever the fluid zone intersects emergent landforms or shorelines. Moreover, when no other conditions are constraining, the segment boundary should be oriented along readily identifiable landmarks.

- (1.2) Submerged reefs and shoals should form a boundary between segments, even when substantial flow over the shoal occurs, because the presence of the shoal will affect detention and circulation both upcurrent and downcurrent.

- (1.3) Channels frequently differ in water quality from the open, shallow bay, and can act as a preferential conduit for flow. Therefore a channel should be included well within the interior of a segment, or be itself an independent segment. Because channels are well-marked by navigation aids, there is frequently a spatial bias of sampling in channels, which should be considered in defining segments that contain channels.

- (1.4) Inlets typically are zones of strong currents. When the inlet is of limited spatial extent, e.g. the entrances to Oso Bay or to Nueces Bay, it is best to separate the zones on either side of an inlet as different segments with the inlet serving as the boundary. When the inlet has considerable spatial extent (or is the site of an extensive base of observations), such as Aransas Pass from the jetties to Harbor Island, the inlet zone should be delineated and identified as a separate segment.

(2) Hydrography:

(2.1) Horizontal gradients in water density, as indicated by temperature and salinity, should be used to define segment distributions, with the zone of shallow (or zero) horizontal gradient lying within the segment interior, and the zone of steep gradient lying on the boundary. Because of the extreme variability of salinity, this definition may have to apply to average or long-term distributions.

(2.2) Any zones of systematic density stratification should be segregated from those of zero or unsystematic stratification. As with (2.1), because of the extreme variability of salinity, this may be a condition to apply to long-term mean density distributions.

(2.3) Current structure, especially current shears, should be employed, when the data are available, to define circulation patterns; generally segment boundaries should lie either parallel to or orthogonal to current trajectories, and should not be oblique.

(2.4) Tidal variation in an estuary can affect water mass retention and water-quality differences. If data permit, regions of particularly dramatic changes in tidal range, and conversions from progressive- to standing-wave properties should be identified and used as a basis for segmentation.

(2.5) Tidal current trajectories are a special case of (2.3) that become especially important in defining segments which contain or are adjacent to inlets or tidal conduits. The complex shoals and channels in the Redfish Bay-Harbor Island area are a good example.

(2.6) Fetch under dominant wind regimes can govern regions of the bay which are well-mixed and those that are not, due to the importance of wind-driven waves in effecting mixing. A subtler effect may be the generation of large-scale wind-driven gyres, but on the Texas coast definitive data on these forms of circulation are lacking.

(2.7) Turbulence derived from bed roughness is an important source of mixing and dispersion, and therefore can be important in delineating areas of differing mixing intensities. To the extent that information exists on bedforms, this should be incorporated into the segment definition.

(2.8) Inflows are a prominent source of systematic (throughflow) currents, and under sufficiently high flows can lead to extensive water-mass replacement. Segment boundaries should therefore take into account the normal region of influence (i.e., the outflow plume) from a point inflow. For large-scale inflows, such as the Nueces River, this is obviously a feature that will vary with river hydrograph, and at times encompass all of Nueces Bay or even Corpus Christi Bay proper, so some judgement may be required. Another type of inflow to consider is the large-volume discharge from an outfall, e.g. the cooling water return of a power plant.

While these hydrographic principles can be articulated, the fact is that the data base upon which these kinds of decisions must rely is usually lacking. Indeed, many of the hydrographic judgements must revert to morphological considerations or to the application of fluid-dynamics intuition.

Water quality is in fact a suite of parameters, each of which is subject to its own complex of sources and sinks, and kinetic processes. To a varying degree, however, transport processes underlie all of these, so the hydrographic properties enumerated above form a set of minimal criteria for segmentation. In addition, water quality segmentation must also reflect the following properties specific to water quality constituents:

(3) Water quality:

(3.1) Regions of homogeneity are one of the most important factors in the definition of segments. Ideally, a segment should encompass a region which is largely homogeneous in water quality. At the same time, water quality parameters in the real world are extremely variable and rarely homogeneous (the term implying a certain threshold of statistical variability which is deemed acceptable). Segment definition is based first upon relative differences in water quality—some regions having a greater tendency toward homogeneity than others—and second upon the kinds of transport and mixing processes that would tend to promote homogeneity.

(3.2) Regions of steep gradients, in contrast, should be the defining property for a boundary between segments. As with (3.1), this is a relative measure which may be frequently belied by data, depending upon external conditions, and must be supplemented by identifying the kinds of transport and mixing processes that would tend to promote steep gradients.

(3.3) Proximity to loads should be considered in defining water quality segments, since this would entail a large-scale difference in water quality that is superposed upon whatever ambient mixing processes are operative.

(3.4) Systematic degradation of a region of the bay, as exhibited directly in trends of water quality or indirectly in anthropogenic influences, is sufficient reason to segregate an area as a specific segment. Reaches of the Corpus Christi Ship Channel and Inner Harbor, and some of the secondary bays, such as Oso, are good examples. This criterion is also related to (3.3), in that proximity to new discharges may be sufficiently compelling to anticipate degradation of water quality.

(3.5) Finally, any region in which there is a systematic trend toward degradation, even though the water quality indicators may still lie within the normal or "healthy" range may be beneficially monitored by being defined as a separate segment. This is a more subtle differentiation of the same philosophy expressed in (3.4).

Ideally, a separate segmentation would be defined for each water-quality parameter, e.g. dissolved oxygen, sediment mercury, BOD, but such an approach would be manifestly unworkable, therefore the definition of segments needs to consider the principal water quality parameters taken collectively.

The application of criteria (1), (2), and (3) without further constraints would result in a veritable plethora of segments. From the opposite direction, we wish to minimize the number of segments in order to: (i) maximize the number of data points per segment, hence the statistical strength of the conclusions, (ii) improve the conceptual value of the analysis, by presenting the results for as few, large segments of the bay as possible (a reflection of the poor ability of the human mind to assimilate numerous facts simultaneously, requiring some degree of predigestion). There is an additional practical criterion lurking here as well, *viz.* to decrease the effort of analysis, which will proceed on a segment-by-segment basis and therefore is proportional to the number of such segments, but this is probably subsumed within (ii). This criterion can be expressed as follows:

(4) Maximal spatial aggregation:

(4.1) Definition of segments should be cognizant of the conceptual value of organizing the system into a small number of quasi-autonomous regions. Further, the boundaries of these regions should correspond to natural physiographic boundaries and be defined by well-established, easily determined landmarks or landforms.

(4.2) Dimension of a segment should take into account the minimum number of data points within the segment required to characterize a spatially representative value. This would be based upon the distribution of historical sampling stations in the region and typical variability in water quality.

(4.3) Dimension of a segment should also consider the minimum number of data points over time needed to resolve principal temporal variability, and the available period of record at the established monitoring stations.

(4.4) One element of establishing acceptable spatial dimensions is the intrinsic variability at a given point in water quality versus variability across the segment area.

(4.5) The segmentation scheme should be comprised of non-overlapping segments, so that every point in the watercourse falls within a unique segment.

Criterion (4.1) appears different in character from the other three, in that it is more qualitative and can be applied from a purely morphological viewpoint, while (4.2)-(4.4) have a strong statistical flavor, and would require a fair data base for their application. Implicit in (4.1) is the concept of areal scale of depiction, which is not an absolute measure but is, rather, at least partially determined by the objectives of the analysis in which the segmentation is to be employed. For some purposes, a rather gross segmentation might appear satisfactory. The extreme example would be analyzing bay-wide parameters, in which the entire bay is regarded, in effect, as one segment. This sort of analysis is done, for example, when bay-wide water budgets are carried, in tidal prism analyses, and in computing bay-wide salinities (cf. the annual reports on bay salinities of the Texas Parks and Wildlife Department). Upon closer analysis, however, these "global" depictions of the bay really amount to accepting a rather large statistical variance in the answers. (Some users of such a gross approach may, of course, be unaware of the enthymeme.) The key point is that the intended spatial scale of the analysis, determined by the objectives of the analysis, in effect imposes a level of statistical variance—a confidence level—on the results.

Criterion (4.5) is not required by the statistical methods of data analysis. Indeed, for scientific purposes, it is perfectly acceptable to have overlapping segments, with sampling stations counted among the aggregated data for more than one region. On the other hand, by requiring disjoint segments, it then becomes possible to carry out a census of data availability, and to avoid problems of weighting of measurements, due to the same measurement being counted in more than one aggregation. Further, the administrative function of segmentation would be greatly complicated by overlapping segment definitions. (In the Corpus Christi Bay segmentation, this requirement is deliberately violated for the Gulf of Mexico hydrographic segments, see Section 3.3.3.)

The definiteness of these criteria is somewhat meretricious, in that the information base for their quantitative application is not extant. This is especially true for the statistical measures underlying (4). For most parameters, these measures are *ab initio* unknown and in any event relatively spongy, varying with intended purpose of the analysis and the cultural bias of the investigator, and varying over time. However, their enumeration serves the good purpose of providing an objective set of criteria that can be applied intuitively, and on the basis of gross characteristics of the bay and past experience with data from the bay. Because these criteria are rather intuitive, the present TNRCC WQ Segmentation more or less conforms to them.

The notion of a scale of analysis emerged in the formulation of criteria. This aspect of segmentation cannot be overemphasized. Underlying any segmentation scheme is a dominant spatial scale of analysis, which carries with it an associated level of confidence one is willing to accept in the aggregation of samples over a region of the estuary. For someone studying the variation of water quality in Corpus Christi Bay on a scale of tens of kilometres, it is appropriate to depict Nueces Bay as one or two segments. Another researcher with the different purpose of studying the kinetics of a constituent within Nueces Bay itself would find this scale of representation much too coarse, and would employ a much more refined spatial segmentation. Either level of segmentation would be inappropriate and unworkable for the other's purpose. (Note that the use of field data from a network of stations implicitly assumes a segmentation, in that each sampling station is presumed to represent water quality over some extended area in which the station is located.)

3.3 Project segmentations of Corpus Christi Bay

3.3.1 Numerical schema of segments

Not only must we define a system of segmentation for analysis of water quality, we must devise a means by which this segmentation may be imposed in automatic data processing. That is, there must be a computational means for determining into which segment a sampling station would fall. The method adopted here is to define each segment as the union of quadrilaterals encompassing the portion of the watercourse defined to lie within that segment. The corners of each quadrilateral are given by latitude/longitude pairs, and an algorithm was developed which determines whether a given station, specified by its latitude-longitude coordinates, lies within the quadrilateral. We note that (1) these quadrilaterals are not parallelograms, but can be any four-sided figure distorted as necessary to conform to the shape of the watercourse, (2) only the watercourse is considered to be within the segment, even though the quadrilateral may cover substantial land area as well, (3) a series of quadrilaterals taken together can better approximate complicated geometry. The last is the reason that a conjunction of such quadrilaterals is used. Because many segments, especially TNRCC segments, have complex shapes, a single quadrilateral was not practical.

Figure 3-1 shows an example of the depiction of segments by quadrilaterals, Segment 2472, Copano Bay. The main body is one segment and is adequately depicted by a single quadrilateral, as indicated. Segment 2003 is Aransas Creek Tidal from its mouth at Copano Bay to approximately Rincon Bend. Segment 2001, Mission River Tidal, extends northward from Copano Bay, and is defined by a large elongated quadrilateral. The key sides of both of these quadrilaterals are those that intersect the watercourse at the boundaries of the segments with Copano Bay. Placement of the other corners is arbitrary and is adjusted to optimize the fit and to simplify the corner coordinates. Nueces Bay, Segment 2482, is represented by three quadrilaterals, as shown in Fig. 3-2.

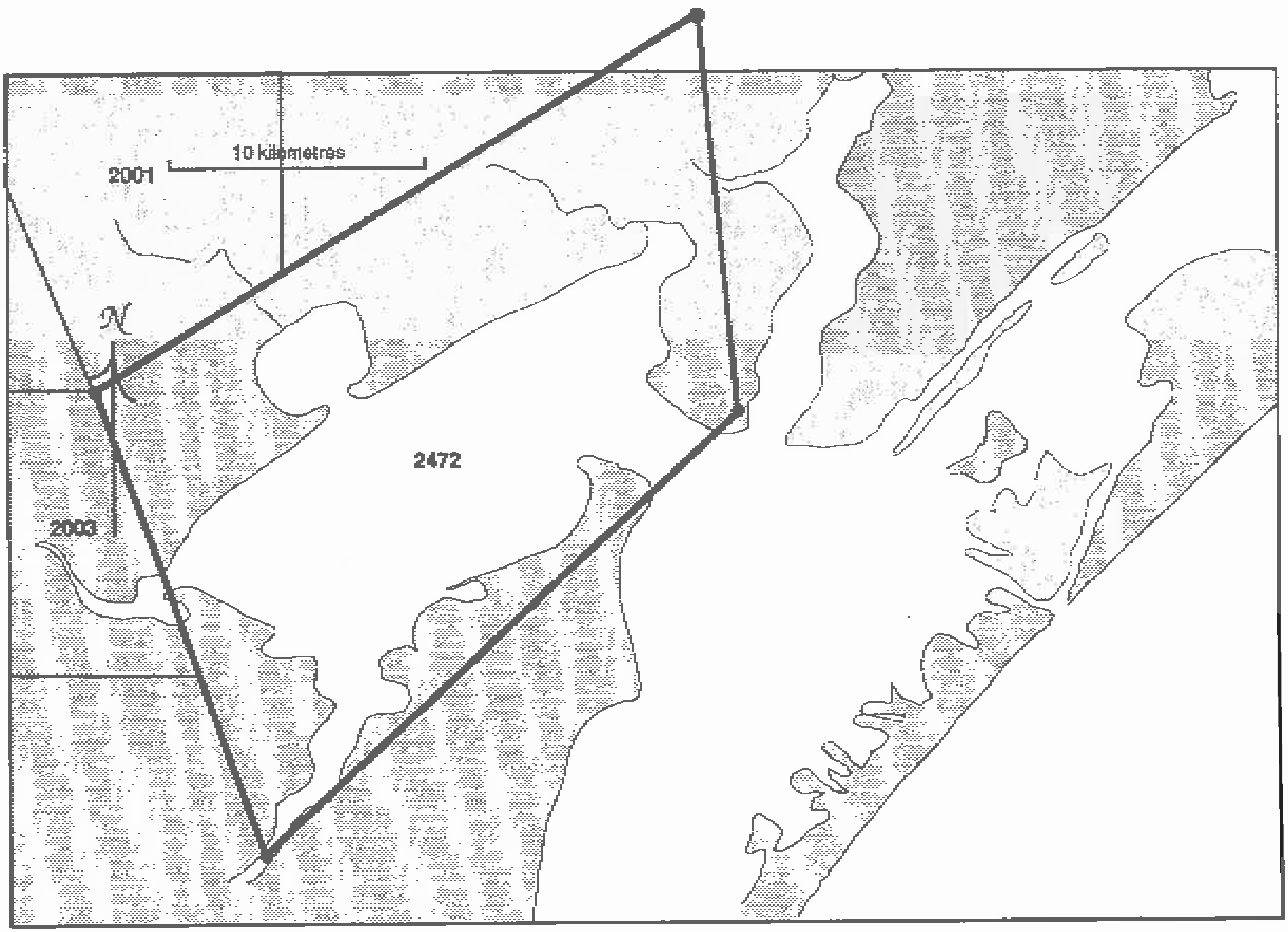


Fig. 3-1. Quadrilateral definition of TNRCC segments in Copano Bay

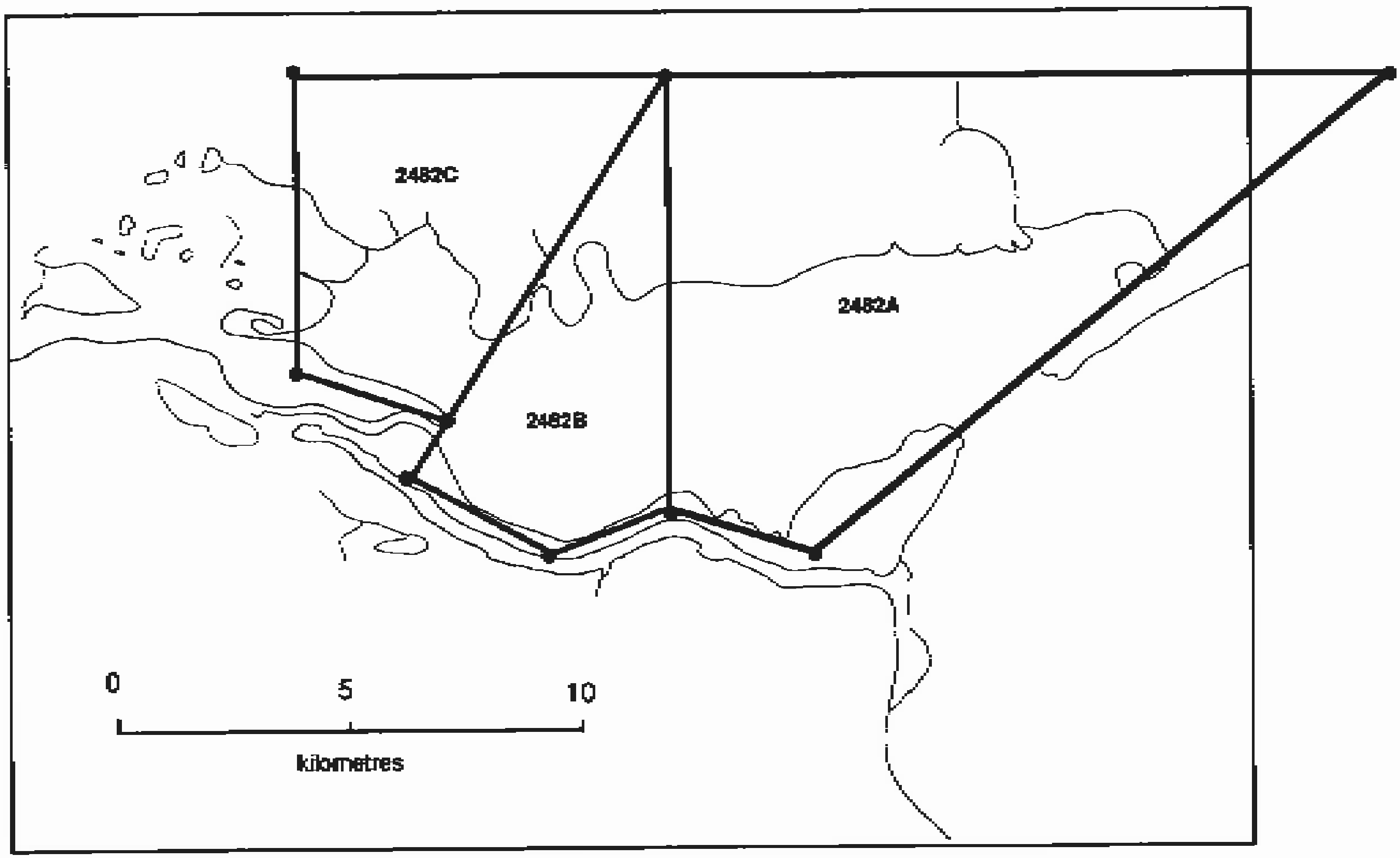


Figure 3-2. Quadrilateral definition of Segment 2482, Nueces Bay

In addition to providing a quantitative mechanism for processing large data bases, the quadrilateral depiction of segments has another benefit: it is a means of precisely and quantitatively defining the boundaries of a segment. The nearly universal practice of using small-scale maps to depict segment locations is a continuing problem, especially in open waterbodies where there is no obvious morphological boundary, as these maps are subject to drafting and printing errors, and introduce uncertainty in the precise location of boundaries. With a quadrilateral-based definition, the quadrilateral corners can be specified to whatever accuracy is needed, and become an unequivocal means of communicating segment boundaries. While in this report, we will display maps of the hydrographical segments employed, it should be understood that these are graphic vehicles only, and the segments themselves are precisely defined by the applicable quadrilaterals.

3.3.2 Texas Natural Resource Conservation Commission Water Quality Segments

As noted earlier, the Texas Natural Resource Conservation Commission system of segmentation forms the basis for water management in the state. The Corpus Christi Bay estuarine system *per se* is represented in this system by sixteen (16) segments, ten (190 in the open bays (including Oso), five (5) in the tidal tributaries (including the Inner Harbor), and one in the Gulf of Mexico. These are summarized in Table 3-1. The spatial scope of this project was extended to include three designated freshwater segments, seven undesignated (both fresh and tidal), and one comprising a portion of the next bay system (San Antonio), added to facilitate the State's assessment of the coastal basin. The undesignated tributaries are administratively taken to be part of the segment into which they conflow, but were added as independent watercourse segments for this study. (The division into tidal and above-tidal is ours.) Those TNRCC segments within and in immediate proximity to the principal components of the Corpus Christi Bay study area are depicted on Figs. 3-3 through 3-7. The boundaries for the open-bay segments are not well-defined and are established qualitatively (i.e., by approximate location on crude maps, e.g. TWC, 1990). Thus far, this has not presented an administrative problem, because the routine monitoring stations of TNRCC generally are placed well in the interior of these segments.

The quadrilaterals used by this project to define the TNRCC Water Quality Segments are given in Table A-4 in the Appendix.

3.3.3 Hydrographic segmentation

It was necessary to formulate a segmentation system suitable for use in this project. The primary purpose of the segmentation is analytical, i.e., for data aggregation by area of the bay, to support statistical and trend analyses. For the establishment of general levels and trends in water quality, the spatial resolution

Table 3-1
Texas Natural Resource Conservation Commission
Water Quality Segments
for CCBNEP Study Area

Segment	Name
Estuarine & marine segments	
2001	Mission River Tidal
2003	Aransas River Tidal
2101	Nueces River Tidal
2203*	Petronila Creek Tidal
2462*	SW San Antonio Bay & Hynes Bay
2463	Mesquite & Ayres Bays
2471	Aransas Bay (including Lydia Ann Channel)
2472	Copano Bay
2473	St. Charles Bay
2481	Corpus Christi Bay
2482	Nueces Bay
2483	Redfish Bay
2484	Corpus Christi Inner Harbor
2485	Oso Bay
2491	Upper Laguna Madre, JFK Causeway to Yarborough Pass
2492	Baffin Bay
2501	Gulf of Mexico nearshore
Freshwater segments	
2002*	Mission River above tidal
2004*	Aransas River above tidal
2204*	Petronila Creek above tidal

Undesignated segments, estuarine & freshwater

Copano Creek*	Oso Creek*
Los Olmos Creek*	Oso Creek Tidal*
Los Olmos Creek Tidal*	San Fernando Creek*
Poesta Creek, upstream from Aransas River to Beeville*	

* added to segments specified in project Scope of Work

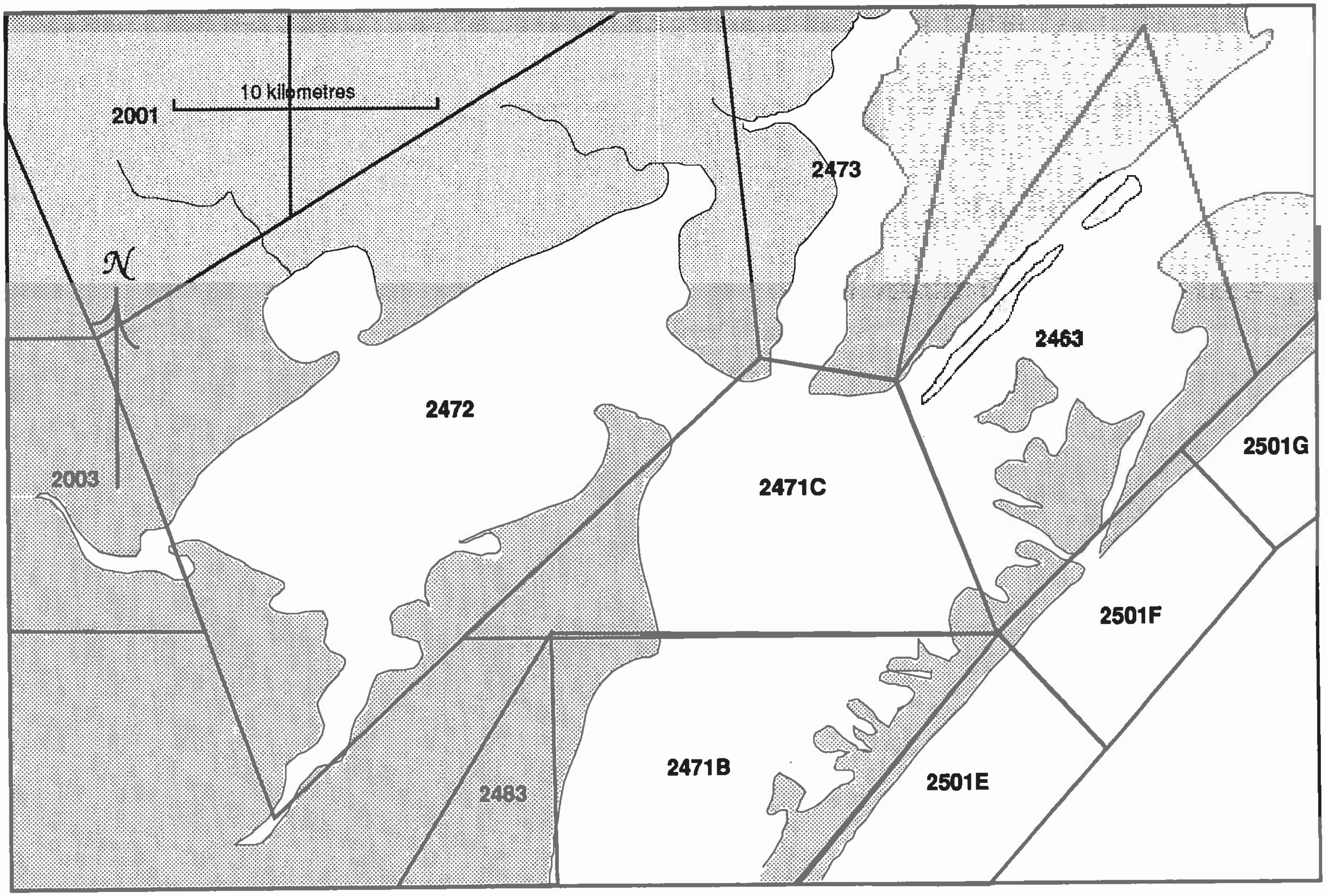


Figure 3-3. Quadrilateral definition of TNRCC segments in Aransas Bay area

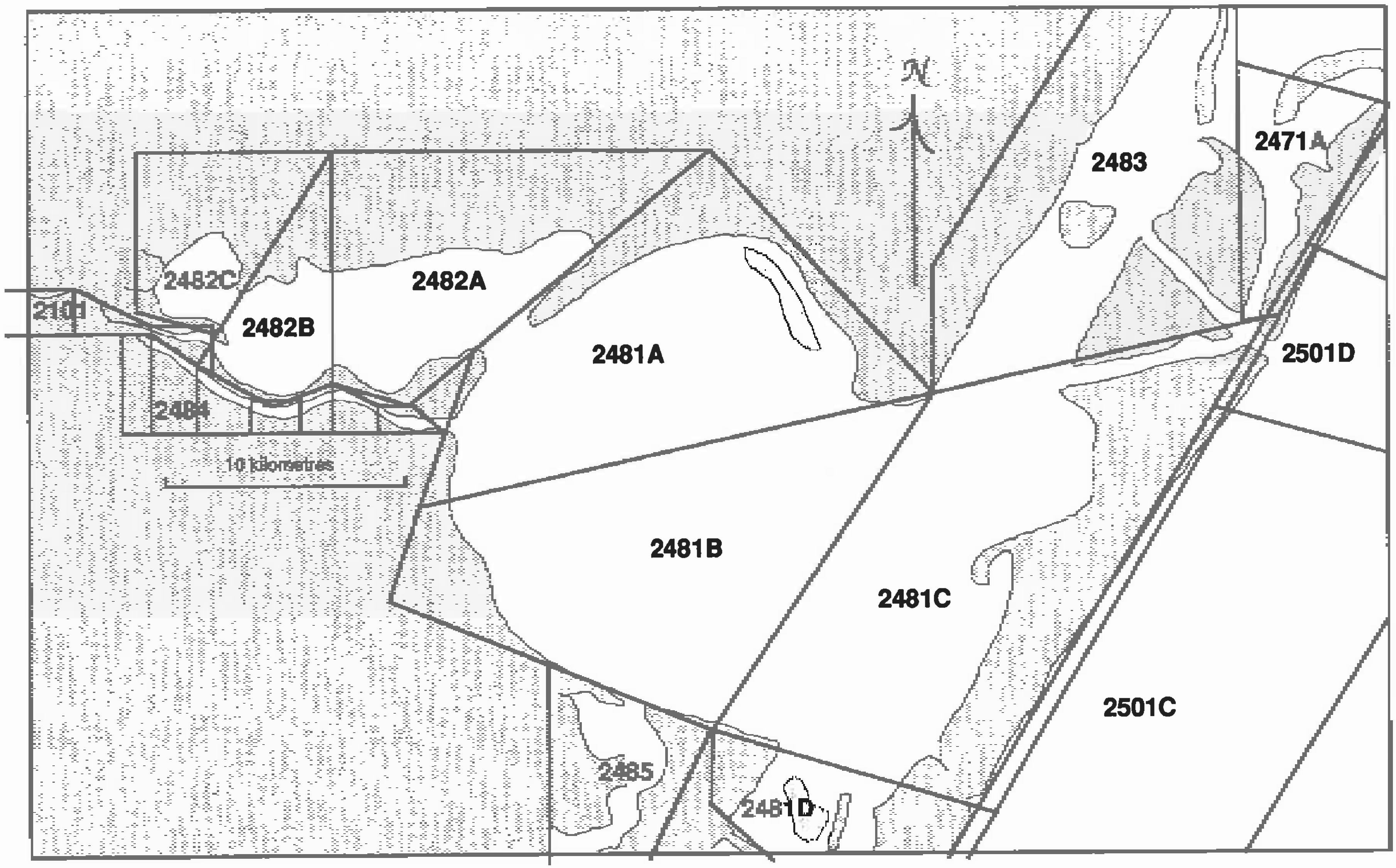


Figure 3-4. Quadrilateral definition of TNRCC segments in Corpus Christi Bay area

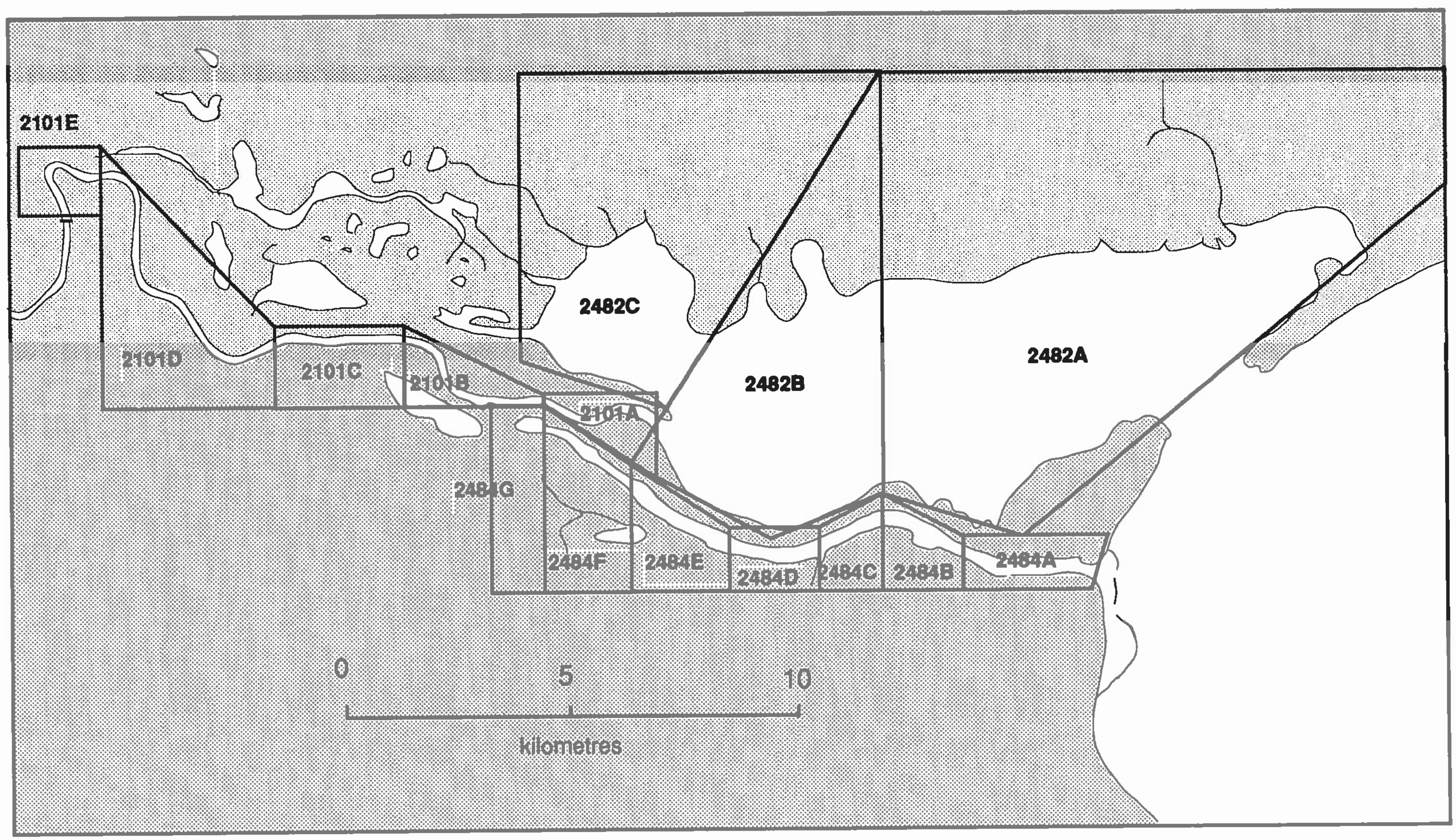


Figure 3-5. Quadrilateral definition of TNRCC segments in Nueces Bay area

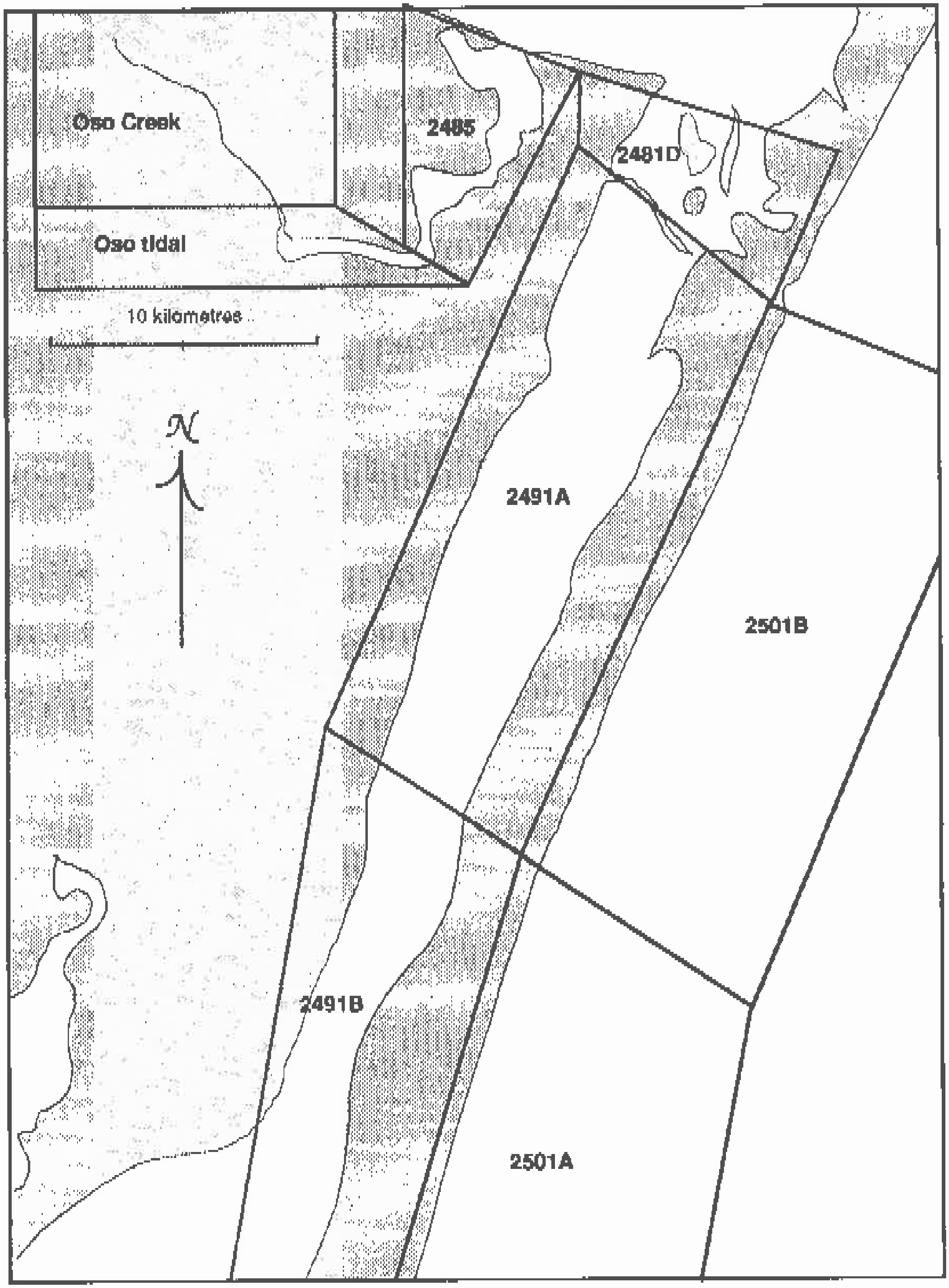


Figure 3-6. Quadrilateral delineation of TNRCC segments in Upper Laguna Madre

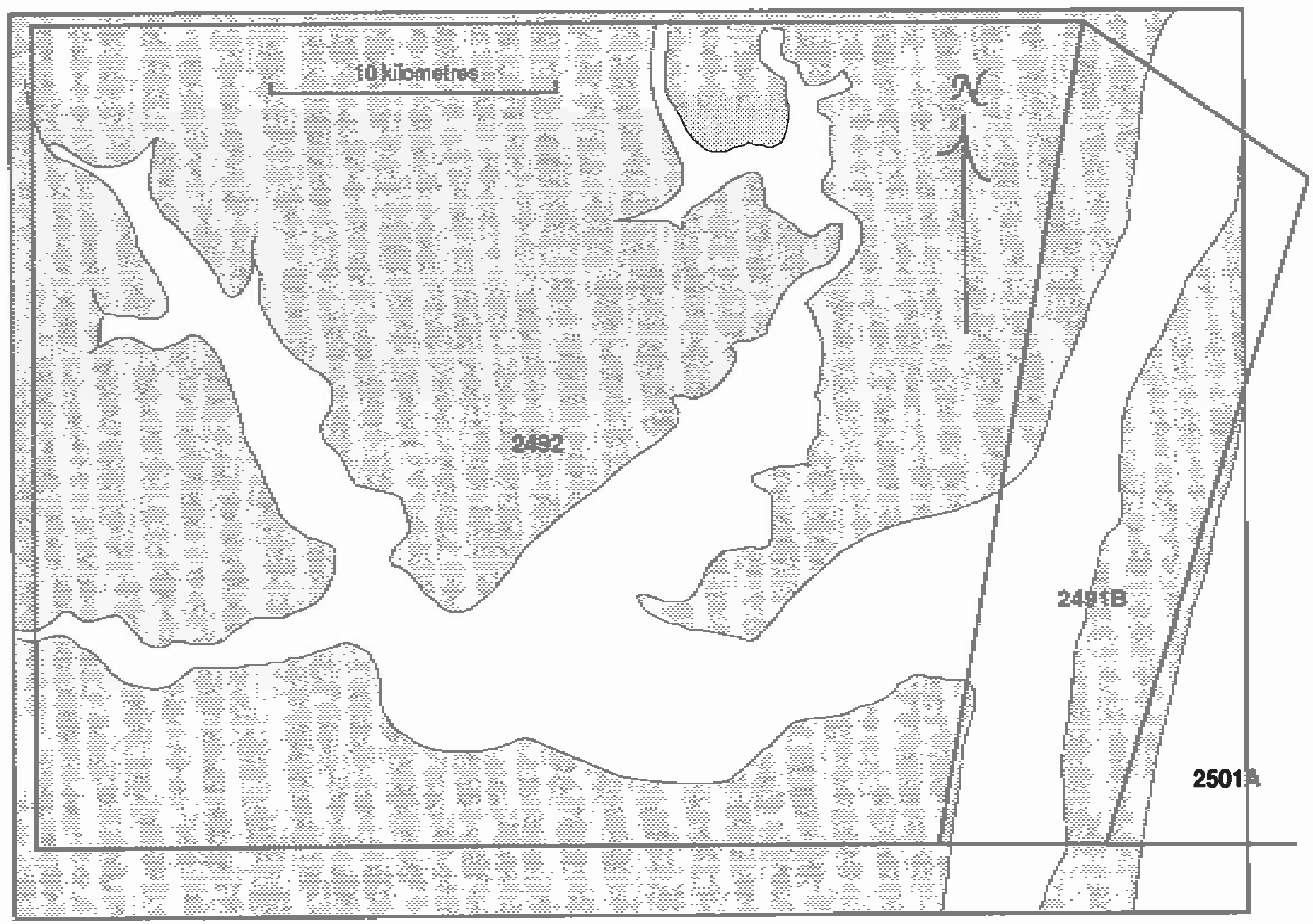


Figure 3-7. Quadrilateral definition of TNRCC segments in Baffin Bay

needed to be on the order of 5-10 kilometres, except when hydrographic properties demand a smaller scale. From a practical standpoint, application of the principles enumerated in Section 3.2 meant basing the segmentation largely upon the hydrographic criteria (1) and (2) of Section 3.2, along with the data management criteria (4). This segmentation considered the criteria (3) to the extent that experience and data on specific water quality variables permitted.

The final hydrographic segmentation for the study area is depicted in Figs. 3-8 through 3-12. This segmentation was limited only to the estuarine and marine portion of the study area. Those segments lying upstream from the areas shown on Figs. 3-8 through 3-12 were considered to be adequately depicted by the TNRCC segmentation analyses. It will be noted that generally the hydrographic segments (HS) of Figs. 3-8 through 3-12 are considerably smaller than those of the TNRCC, so a finer level of spatial analysis is permitted, especially in the open bay areas. There are a total of 178 hydrographic segments, distributed as indicated in Table 3-2. The coordinates precisely defining the locations of these hydrographic quadrilaterals are given in Table A-5 of the Appendix.

These hydrographic segments formed the fundamental organizational units for the water quality and sediment data in the present project. Some particular features of this segmentation warrant mention. The Corpus Christi Ship Channel in the open bay occupies its own segments, a narrow strip of approximately 2 km width centered on the dredged channel. Similarly, the La Quinta Channel and prominent reaches of the Gulf Intracoastal Waterway (GIWW) are also embedded within narrow segments. This is due to the peculiar hydrodynamics of salinity intrusion and increased tidal response dictated by the deeper water, and also due to the isolating effect of dredge disposal areas on the lateral boundaries of these channels. One rather odd-appearing segment NB7 encloses the return from a major power plant. The orientation of the segments in Corpus Christi Bay follow the curvature of the bay with the narrow dimension perpendicular to the shoreline. This is in anticipation of gradients in quality produced by runoff and discharges from the bay periphery. The boundaries of several of the segments are dictated by reefs or other bathymetric features. For example, CP03 in Copano Bay is bounded on the east by Shellbank Reef, CP06 is bounded on the west by Copano Reef and on the east by Lap Reef, and RB1 through RB9 encompass the complicated bar, channel and shoal complex of Redfish Bay. ND1 through ND4 are the Nueces marsh/delta area, and NR1 through NR5 comprise the channel of the Nueces River below Calallen Dam.

In the nearshore Gulf of Mexico there are no physiographic boundaries other than the barrier island complex that would control segmentation. Segmentation here was based upon two features: water depth and proximity to a tidal inlet. Fig. 3-13 shows a typical depth profile extending from the barrier island out into the Gulf of Mexico. There are two prominent breaks in the slope of the seabed, one at approximately 10 m (5 fathoms) and one at approximately 20 m (10 fathoms). Hydrographic segments were defined to extend between these two breakpoints.

Table 3-2
Distribution of Hydrographic Segments
in principal subdivisions of the study area

system	*segments*
Mesquite/Ayres Bay	9
Aransas-Copano Bay	36
Redfish Bay and Aransas Pass	16
Corpus Christi Bay proper*	42
Nueces Bay and Nueces River tidal reach	18
Inner Harbor	7
Upper Laguna Madre	23
Baffin Bay	9
Gulf of Mexico nearshore	18

* including Oso and La Quinta Channel

The innermost zone, from the shore out to the 10 m contour, is one of relatively steep slope, and extends about 2 km into the Gulf, encompassing the surf zone and the larger zone in which substantial refraction of the longer-length waves occurs. The outer zone, from the 10 m to the 20 m contours, extends out about 10 km into the Gulf. In the vicinity of tidal inlets, the water quality can be expected to be influenced by exchange with the adjacent estuary. Therefore, hydrographic segments were defined to correspond to the main extant and historical passes. All together, the Gulf nearshore is divided into nine (9) sections, each of which has an inner and an outer segment, for a total of 18 hydrographic segments. These are shown schematically in Fig. 3-14. A close inspection of these segments and those defined for the estuarine areas will disclose that the inlet segments overlap to a certain extent. Therefore sampling stations located within the inlet will be captured in both the estuarine segment and the Gulf of Mexico segment. This is deliberate. These transitional stations are considered to be indicative of both the nearshore Gulf of Mexico environment as well as that within the inlet itself, and therefore need to be included in both data subsets. (This is discussed further in Chapter 6.)

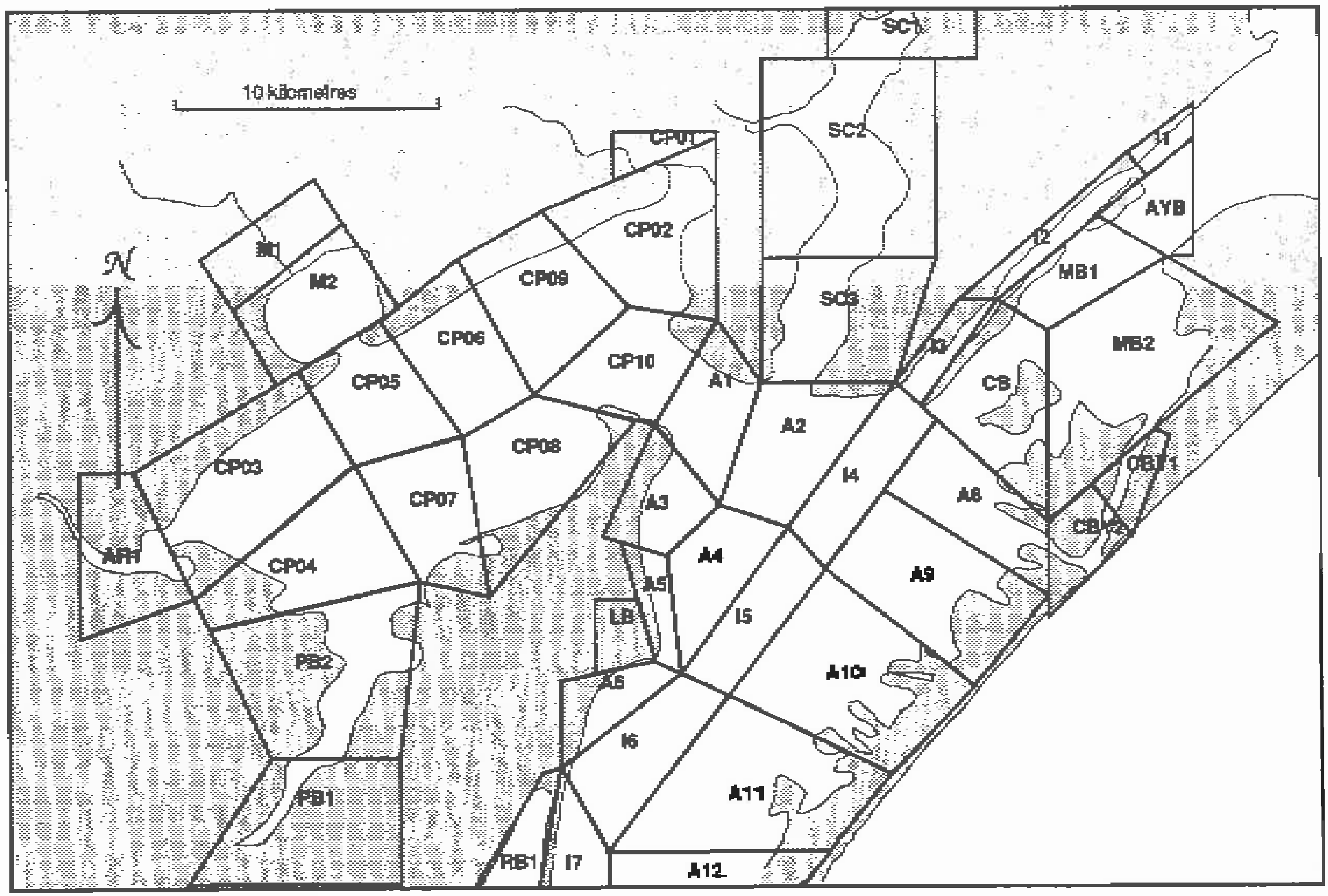


Figure 3-8. Hydrographic areas for Aransas-Copano system

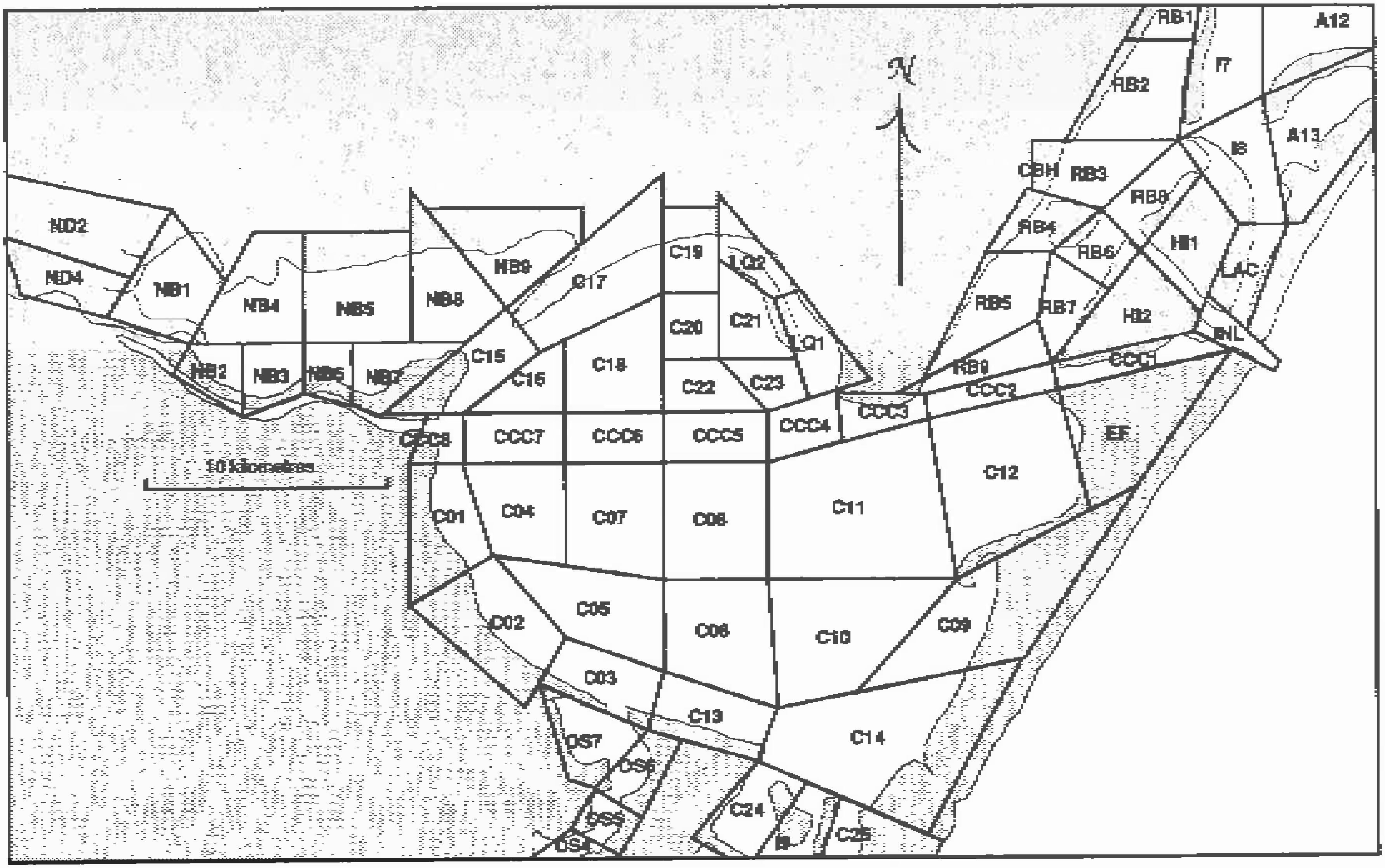


Figure 3-9. Hydrographic areas for Corpus Christi Bay

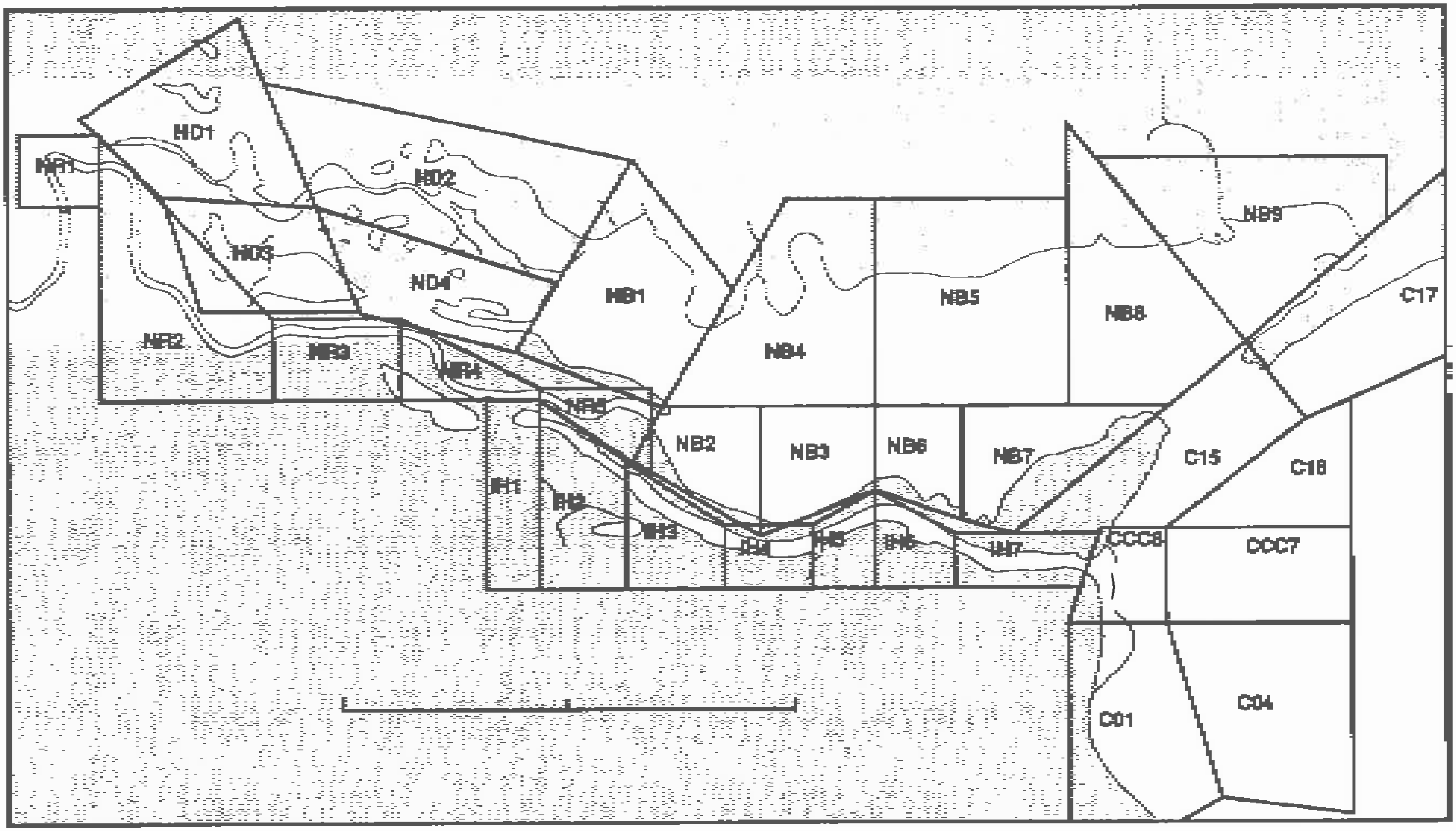


Figure 3-10. Hydrographic areas for Nueces Bay region

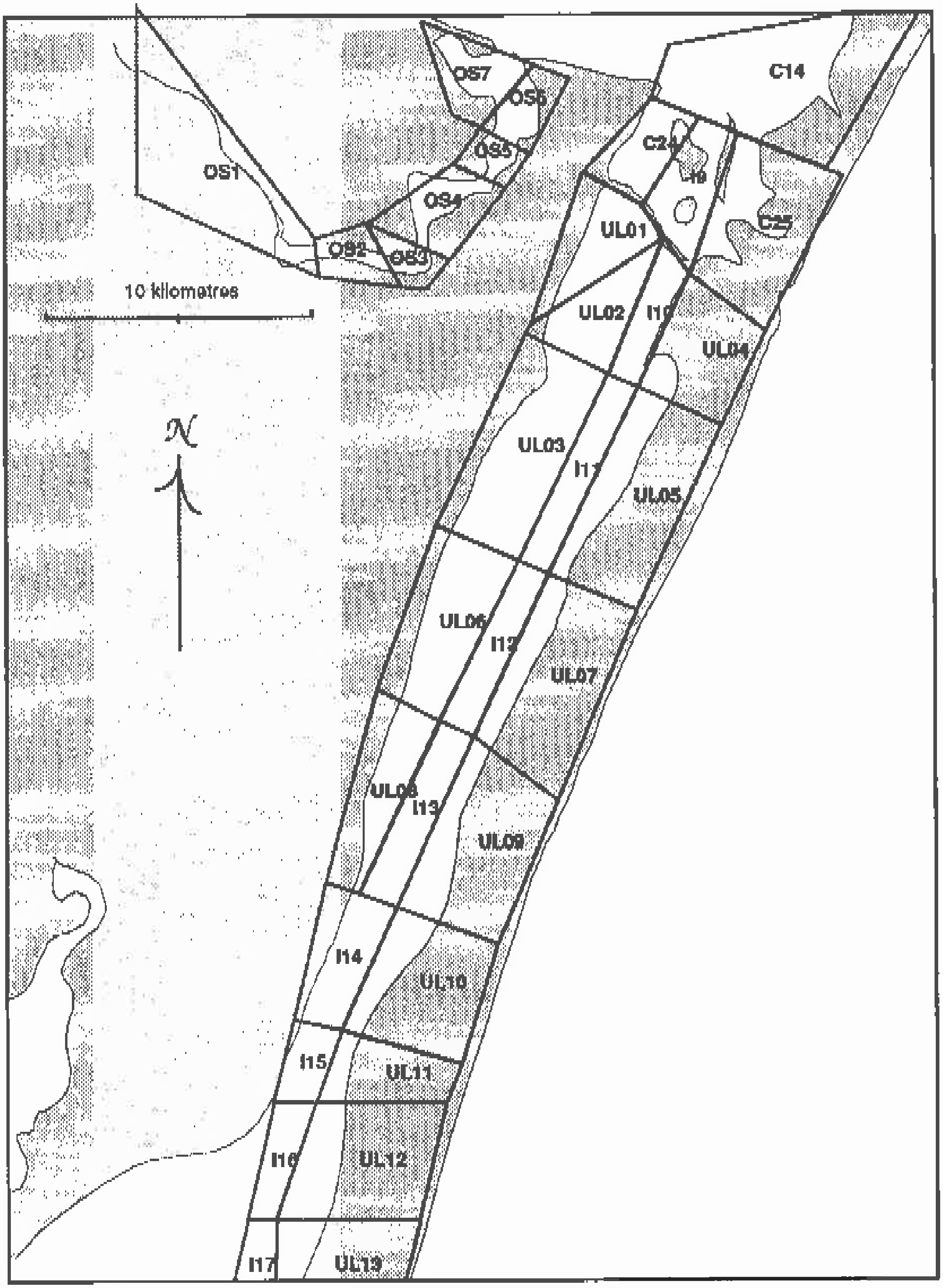


Figure 3-11. Hydrographic areas for Upper Laguna Madre and Oso Bay

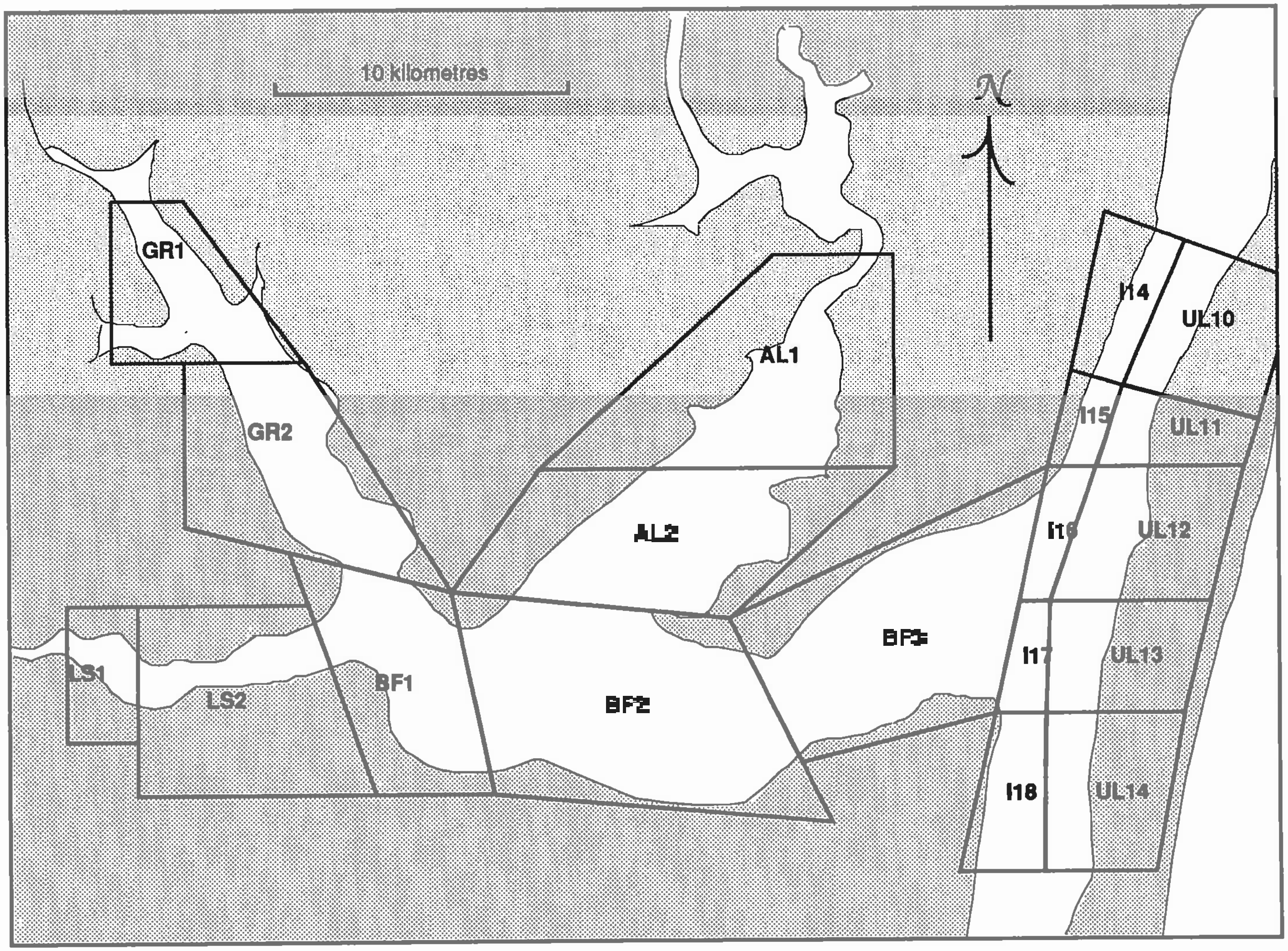


Figure 3-12. Hydrographic areas for Baffin Bay and adjacent Laguna Madre

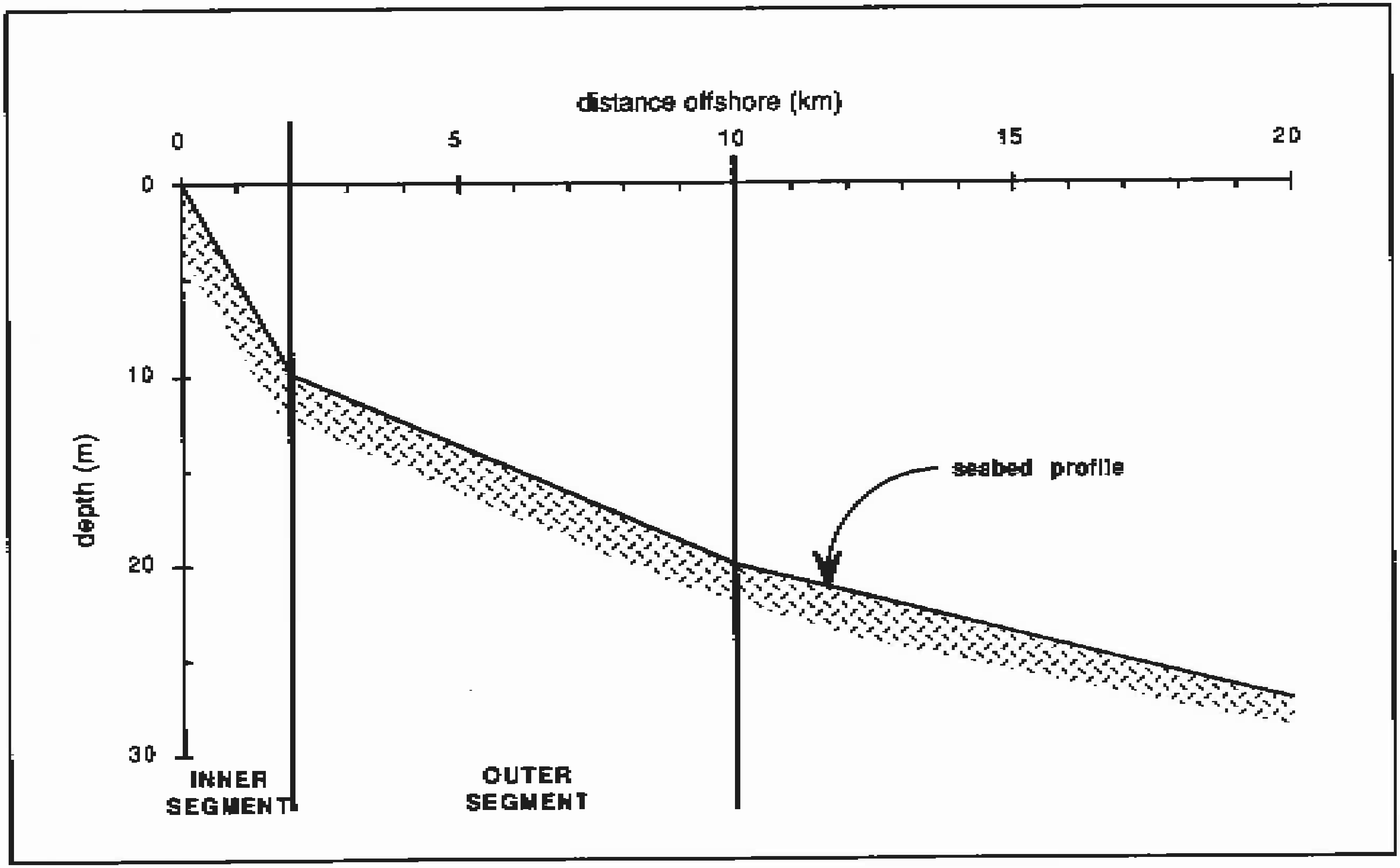


Figure 3-13. Definition of hydrographic segmentation in Gulf nearshore.

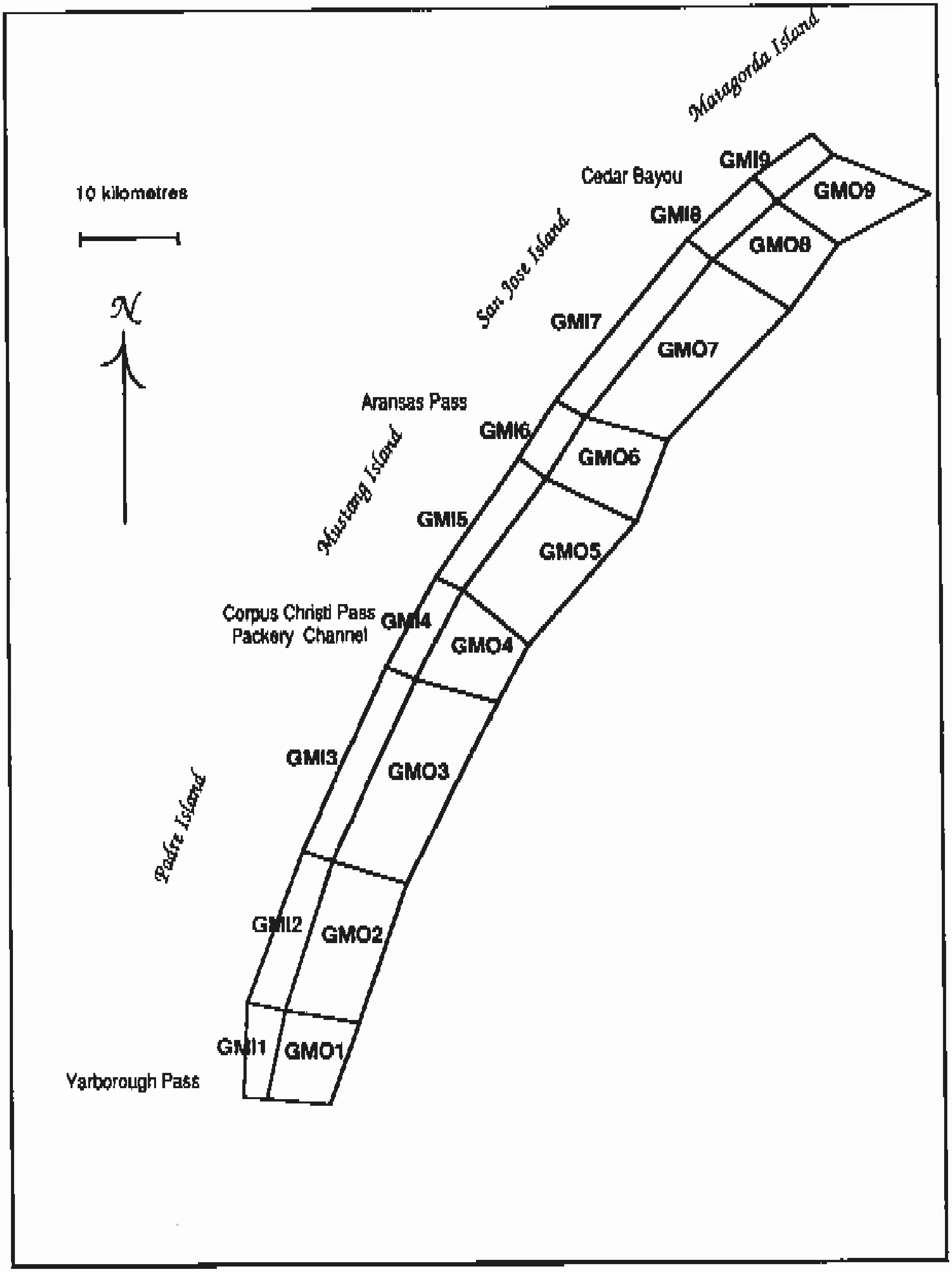


Figure 3-14. Hydrographic areas for Gulf of Mexico nearshore

4. DATA COMPILATION AND PROCESSING PROCEDURES

A major task of this study was the acquisition, compilation and digitization of historical data relating to the quality of water and sediment within the Corpus Christi Bay system. The data acquired in this project can be broadly categorized as digital format and hard-copy format. The former refers to any medium capable of manipulation on the digital computer, e.g. magnetic tape, floppy disks, CD's, Internet-accessible digital data files, etc. The latter refers to field sheets, typewritten tabulations, and (sadly) computer printout from digital files that no longer exist. In the case of the hard-copy data, all of the significant data sets, and most of the insignificant (that we had access to), were keyboarded as a part of this project effort. This proved to be an extensive process, undertaken by the employ of a welter of data-entry gnomes who hammered away at the data sets over a period of months. It is probably not inaccurate to observe that the probability of marshalling this kind of data-entry effort in the future is unlikely, so certainly one of the major products of this project is the digital data base itself, which is described below. The further analysis of these data requires their conversion, combination and transformation in various ways, all of which can circumscribe the interpretation of the data. The general procedures used in this project are outlined here.

4.1 Data Set Construction

Because the data in this compilation was to be analyzed in later tasks of the project, part of the effort was invested in integrating the data into a computer-manipulable data base. The same data format was used that was designed for the Galveston Bay National Estuary Program project (Ward and Armstrong, 1992a), to facilitate transfer and use of the data by other workers. In designing this format, emphasis was placed on data structure that is transferrable and manipulable via microcomputers (especially PC's and workstations), i.e. compact ASCII files. It may be noted that Tetra Tech (1987) recommends a specific hierarchical format for NEP data sets. While certain aspects of this format are satisfactory—or at least workable—the recommendation suffers from two great deficiencies: (1) the data structure contains numerous redundant fields, which will greatly expand the storage requirements for a data set, (2) the structure is specific to the statistical software package SAS, which will necessitate either that software for its use or specific codes for conversion. (SAS has a fairly robust repertoire of input formats, so this is not so serious a limitation as for some other softwares, especially the data base managers, but the SAS requirement that missing data be represented by a decimal in the otherwise blank field will require pre-processing any ASCII files not observing that convention. It would be straightforward to create a program to read the data sets from this project and generate data files in the format of Tetra Tech, 1987.) Details on the data sets themselves, the formatting of the data base, and related processing information are given in a companion report, Ward and Armstrong (1997a), which is intended to serve also as a User's Guide to the Corpus Christi data base.

One of the principles observed in the construction of the Corpus Christi Bay data base was the maintenance of integrity of the original data from individual surveys. That is, in the compilation of data for a given parameter, say nitrogen series, the coded information included identification of the data source, say TNRCC Statewide Monitoring Network versus Corps of Engineers versus TWDB Coastal Data System, and was input without any modification, including retention of the original units of measurements. While the various data sources were later combined in various ways as a part of different analyses, it is mandatory that the data compilation be capable of separating and identifying, say, nitrogen data from divers agencies, as they may differ in accuracy, methodology and procedure, differences which could become crucial in interpreting apparent trends or in more specialized analyses.

This is one aspect of differentiating the *source* data base from *derivative* data bases. The source data base codifies (in machine format) the original measurements as reported by the originating agency. This data base therefore contains exactly the information in the original: nothing is lost or added. Even an apparently innocuous conversion of measurement units can introduce a distortion. For example, many units carry an implicit level of precision that is modified when converted to another system, such as converting depths in feet to depths in meters. Of course, in adapting the data file to the needs of the project, the source data file may be re-formatted. This might entail re-ordering of the variables, removing unneeded or redundant fields, or re-writing in a more compact format.

An excellent example is use of data from the Texas Natural Resource Conservation Commission Statewide Monitoring Network (SMN) data file. This data was provided as a Internet ftp download of separate records of variable length corresponding to sampling "point-events," e.g. a station-depth-date/time record. The record included a listing of the parameters measured, each concentration preceded by the STORET code identifying the variable. In our processing, data were extracted from this massive file and used to build up ASCII data bases of selected suites of variables that then functioned as the *source* data files for further analysis. These *source* files contain exactly the measurements in the original master ftp files, in the original units: only their format of ordering/storage is altered. One specific re-formatting to which all files were subjected was to be ordered in time; the resultant file we refer to as a *primary* file, but it is nothing more than a chronological source file.

For various analytical purposes, however, these data must be modified, for instance converted to common units, averaged in the vertical, aggregated, or screened out according to some criterion. The data set so processed is a *derivative* data base. Any number of derivative data bases can be created according to the needs of a scientific investigation; it is our opinion, however, that the source data base, once established, should remain inviolate and sacrosanct. Thus the basic approach in this project was to first create the source data base for a given parameter through the data compilation effort. Then various derivative bases were formed to selectively include certain subsets subjected to specific processing.

The goal of this data compilation, simply put, is to create a derivative record of time/space/concentration for each water/sediment-quality variable of concern. That is, each data entry must identify a point in space-time at which the measurement was performed and the associated parameter magnitude. (This in turn introduces a sense of scale, or resolution, dictated by the resolution in the data as well as our conception of space-time variation in water/sediment quality, and underlying the analyses to be performed. This is discussed with respect to resolution and errors below.) Each data record also includes coded information identifying the data source, e.g. TNRCC SMN, Corps of Engineers, or TWDB Bays and Estuaries Program.

Almost all of the data sets include the time of sampling, at least to some resolution. The point in space of sampling is more problematic. Most sampling programs express position by an alphanumeric station name. In order to be able to process the data spatially, this point must be expressed quantitatively. In this project, latitude/longitude coordinates were used to locate the horizontal position of the sample, and depth (i.e., distance below the water surface) to locate the vertical position. The former required precisely plotting the sampling stations from descriptions or from project maps and determining by manual measurement the coordinate positions, which were then keyboarded into a digital data base. This station location data is entered into a separate file, and the horizontal coordinates merged with the measurements at a later stage of the processing into the derivative files.

In a minority of instances, the data-collecting agency or source included latitude/longitude coordinates for the sampling stations. Even in these cases, we plotted the coordinates as provided by the agency and compared them to descriptions of the sample stations or location maps, or, lacking these, to logic as to where the samples were likely (or unlikely) to have been taken. We have encountered numerous errors, and were forced to plot and re-measure many of these, after receiving corrected information from the source. These problems are described in detail on a data source-by-source basis in Ward and Armstrong (1997a).

4.2 Quality Assurance, Reliability and Uncertainty

The limits of resolution of measurements and the associated imprecision, and the extent of infection of a data set with errors contribute a degree of uncertainty to each entry in the data record. The obverse concept is the reliability of that data set for scientific analysis. The need for determining the reliability of historical data and discounting measurements that are judged to be "unreliable" is clearly important. This is recognized by EPA and general methods for accomplishing this are outlined by Tetra Tech (1987). Further, this need was identified specifically in the NEP Scope of Work for this project. It is the PI's conviction that such judgements must be formulated carefully, and the rejection of data be given close consideration.

In data compilation and processing in this study, a major concern was the detection of errors capable of elimination and the quantification of the residual uncertainty in the data. This includes, but is not restricted to, the procedures commonly referred to as Quality Assurance/Quality Control.

4.2.1 Data screening and data-transfer quality assurance

The CCBNEP primary data bases were compiled from various original data sources, some digitally and some manually, and because a transfer of information is involved, there is the possibility of error. Therefore, specific measures were introduced to minimize the occurrence of error, and maximize its detection, as follows:

(1) All data available in machine readable form from an originating agency were obtained, manipulated and entered in that form. Further, intermedia transfers were minimized, i.e., copies were sought as ASCII or WK1 files that could be downloaded by ftp via the Internet or transferred on diskettes.

(2) Data entry by hand employed standardized formats that mimicked the hard-copy sources, and the data entry methods employed standard, simple software, *viz.* EXCEL or LOTUS-123. The data entries were verified by line-by-line comparison of a hardcopy printout of the entry form with the original data source. (*NB*, screen-versus-original comparison is prone to misinterpretation and fatigue, hence the emphasis on comparison of the two hardcopies.) Following the entry and verification steps, the data were scanned and spot checked personally by one of the PI's.

(3) Each new data set that involved a large file of information (and hence especially prone to errors of fatigue or oversight) was subjected to machine screening to verify that the variables lie within expected ranges and exhibit "natural" variability. When aberrancies were detected, the entries were verified against the original source. In many instances this screening detected apparent blunders in the source file itself. These are discussed separately in Ward and Armstrong (1997a). Further, additional steps in the data processing process included various error traps and cross checks, which serve as further error-checking.

Early in the NEP project work on the Galveston Bay Project, i.e. around 1990, an attempt was made to use optical scanners for data transfer. It was quickly learned, by experimentation with several scanners of varying easel dimensions and manufacturer, that the error rate (i.e., character-recognition problems or noise responses) was unacceptably high, even when the source document was clean and high-contrast. Therefore, manual entry had to be pursued. In this project, the possibility of using scanners with OCR software was tried again. Though much progress has been made in five years, we determined that the error rate was still so high that the labor of correcting these errors, combined with the time to carry out the scanning itself, would equal or exceed the effort of simple manual entry of the entire data set. So once again, we opted for the latter.

Particular note should be given the term "mimicked" in (2) above. This is a significant departure of the procedure of this project from that recommended by Tetra Tech (1987), who require that re-formatting into a uniform format, as well as conversion and/or mathematical transformation, be carried out as part of the data-entry process. We believe this strategy is seriously flawed. The entry of thousands of numbers by keyboarding personnel demands maximizing efficiency and accuracy. Any differences between the keyboard format and the hard copy source are an invitation to misinterpretation and transcription mistakes. Further, since keyboarding personnel are rarely equipped to interpret the numbers they are entering, they should not be expected to carry out calculations of any kind, but to simply input what they see. The Tetra Tech procedure, we believe, reduces efficiency and requires an additional level of oversight that could be totally replaced by machine screening. Moreover, we take exception to the philosophy of altering the source data, even by units conversion or rounding, as discussed above, and this is precisely what Tetra Tech recommends.

4.2.2 Data screening and data-base quality

The errors introduced by the data transfer procedures of this project were the simplest to deal with, because their existence (i.e., that they were in fact errors of entry) could be confirmed by comparison with the original source, and corrections could be expediently implemented. The same screening process, i.e. testing for values within "reasonable" bounds (discussed below), spatial continuity (as reflected by simultaneous data from different depths or nearby stations) and temporal continuity (comparison with measurements at the same station before and after the sampling time), occasionally detected aberrant values in the source data files themselves. When possible, we contacted the agency source to verify the reported information. For most of the data files, however, there is no longer an authoritative source with which to compare the reported data: the original field sheets are discarded, or the principal investigator or originating agency is not accessible (or even extant). This forced us to make probability judgements. Consonant with our philosophy of leaving the source data files sacrosanct, "corrections" were introduced into these data files only when the typographical error was "patently obvious." For example, obviously misplaced (or omitted) decimal points, ppm entered instead of ppt (or vice versa), dropped or inverted digits in a date where there are other data from the same sampling run to confirm the date, are regarded as "patently obvious," and represent the limit to which we entered corrections into the source data files. If there is any reasonable possibility that the source data could be entered correctly, or if it is probably wrong but we have no logical, near-certain means of supplying the correct value, then the entry was allowed to stand. Clearly, most apparently aberrant values fell into this category. In the process of creating the derivative data bases later, and certainly in the later analyses, there is the opportunity to reject apparently aberrant data, so leaving such values in the source files causes no harm to the analyses and preserves the integrity of the source data base.

Latitude and longitude coordinates were also subjected to screening. This employed a "range of limits" screen to verify that the positions fell within the broad latitude range of the Corpus Christi Bay study area of 26° 30' to 28° 30' (which helped in identifying wildly incorrect points as well as data from other bay systems that had crept into some of the source files) and a comparison of station descriptions to where the station coordinates actually plotted. Generally, finer corrections were reserved for the derivative data-base screening unless some independent information was available.

Errors in the positions determined by this project proved to be rare, due to the procedures of cross-checking and proofing used during the georeferencing work, and were easily corrected. However, the latitude/longitude coordinates provided by some of the agencies exhibited problems. For the TNRCC SMN data, for example, some stations were obviously wrong. Station 13420 plotted out in a mudflat on Indian Point. Station 13287 plotted in the center of Commercial Street in downtown Aransas Pass. For some of the stations, the description was too vague to unambiguously position the station. The best information available for each station was used for a "best-guess" position on USGS 7.5-minute quads, and copies of the maps were sent to the TNRCC District Office for verification or correction. Jim Bowman of this office helped immeasurably by going through these maps and marking the real locations of his sampling stations. For the Parks and Wildlife Department hydrographic data base, latitude/longitude coordinates were provided by the agency, but with no independent information on station location information. In this data set, the "range of limits" screen disclosed a number of erroneous coordinates (one such point plotted off the coast of North Africa). While the incidence of error is on the order of 0.1%, positions this much in error were obviously due to incorrect digits in the degrees position. Errors in minutes or seconds would not be so easily detected, if detectable at all. This does give an indication of the probable fraction of location errors in this data set due to simple data entry or map-reading errors.

A new source of station location error encountered in this data compilation that was not manifested during the earlier Galveston Bay data compilation is due to increasing use of high-technology positioning systems. Some, like LORAN navigation, have been in use in marine operations for years. Others, like Geographical (a.k.a. Geodetic) Positioning Systems (a.k.a. Satellites), GPS, are a relative newcomer. These are now being embraced by field sampling operations, because of the recent availability of economically priced receivers and processors. These positioning systems as a source of error are addressed in Section 5.3.1 and in more detail in the Data Base Report (Ward and Armstrong, 1997a). In the present context, we note a proclivity to utilize these systems without adequate understanding of their principles of operation and their limitations, to regard the results with unquestioning veneration as absolute, and to employ no backup or verifying independent location data.

4.2.3 Uncertainty measures

The screening procedures outlined in the two preceding sections address data errors of the typographical or "blunder" variety. There remains, of course, a residual error in any set of measurements, deriving from the omnipresent sources of imprecision, inaccuracy and mistakes (including data-entry errors). In this project, data bases for specific variables were created by the combination of data sets from different sources, with differing analytical methodologies, different agency objectives, and differences in field procedures. In order to be able to attach a degree of uncertainty (or its complement, a level of confidence) to such a data set, it is necessary to assess the uncertainty in each of the component data sets, and devise a means of transferring this information to the composite data set. A data user then has the basic information to further determine how the uncertainty is affected by whatever processing of aggregation, units and proxy transformations, and averaging the data may be subject to.

Clearly, the first step is to define carefully and precisely the formulation of uncertainty, i.e. if we denote the measured value of some parameter as

$$v \pm e \qquad (4\text{-}1)$$

the "error" number e must be defined unambiguously and its dependence upon the value of v and other factors carefully specified. Unfortunately, there is terminological chaos in the practice of reporting uncertainty in science and engineering. In the above expression, e may mean the standard deviation of measurements about their mean, the standard error about the "true" value, the tolerance of measurement, the absolute bounds on the range of v, or the magnitude of some fixed multiple of standard deviation or standard error. With careful definition, the measures relative to the mean are usually interconvertible (the exception being the absolute bounds on range). The problem is that in the literature "±e" frequently appears without any associated clarification. Here we will define how the expression is used in this report, and what meaning we will assume when the data source or reference uses but does not define the expression.

In this report, we will employ the error bound e to be the magnitude of the population standard deviation (σ) about a fixed value of the variate. Specifically, for a given measurement procedure under a static set of controls (same concentration, same lab, same personnel, same equipment, same coffee), e is the standard deviation about the known value of the variate for a theoretically limitless set of replications. For practical purposes, we usually have to estimate e by the standard deviation about the mean of the measurements under the same idealized limitless conditions. (The distinction is one of accuracy versus precision.) This standard deviation is estimated in practice by a finite set of measurements: if the set is large then the estimate is good; if the set is small, the sample standard deviation may have to be corrected to estimate the population standard deviation. For many of the trace organics of current concern in water quality, e.g. the priority pollutants, available precision data may be limited to only 3 or 4 replicates for a given set of controls, so the correction may be substantial.

Technically, the correction is the factor $\sqrt{[N/(N-1)]}$; further, other statistical inferences must be altered if the sample standard deviation departs significantly from the population standard deviation. This and related matters are treated in any standard textbook on statistical methods, e.g. Mood (1950), Hamilton (1964).

Standard Methods (APHA, 1971, 1985, 1989) and ASTM Standards (ASTM, 1976, 1980) have historically recommend the use of "±e" as a standard deviation. Unfortunately, these proposals for a uniform reporting of precision compete with practice and intuition in the literature. Many authors use "±e" to specify, in effect, tolerance limits, i.e., the range in which "most" of the measurements fall. "Most" seems to mean substantially more than 95%. Tolerance specification has traditionally assigned a level to e of about 3σ (e.g., Kennedy and Neville, 1976); exactly 3σ implies a 2.7% probability of a measurement with normally-distributed error falling outside the indicated range, while 3.09σ implies exactly a 2% probability of violation. This usage seems to lie much closer to the intuitive connotation of precision expressed as ±e than the use of standard deviation, especially among water resources scientists. Another competing concept is the precision latent in the expression of significance. A measurement reported as 5.36, for example, with no additional qualifiers, implies a tolerance (in the above terminology) of ±0.01 or no worse than ±0.02: by writing the third digit, the author is indicating relatively strong certainty of its significance (EPA,1979). (Mathematicians are generally more fastidious, demanding a tolerance < 0.5 times the last significant unit, e.g. ± .005 in the above example, Scarborough, 1966: then the statement "correct to *n* significant figures" means *correct* to *n* significant figures. Skougstad et al., 1979, state that the last significant digit is the "first doubtful digit," but the meaning of "doubtful" is not elaborated.) Thus, for a worker known to be scrupulous in the expression of significance, some measure of that worker's judgment of precision can be inferred.

The National Institute of Standards and Technology (NIST) has attempted to bring order into chaos by a new (October 1992) policy on expressing measurement uncertainty, based upon recommendations of the International Committee for Weights and Measures (1992). In these authors' view, NIST has furthered the chaos by introducing new, unnecessary terminology:

standard uncertainty	-	an estimated standard deviation
Type A evaluation	-	based on "any valid statistical method"
Type B evaluation	-	based on "scientific judgment using all relevant information"
combined standard uncertainty	-	estimated standard deviation from the standard uncertainties of both Type A and Type B
expanded uncertainty	-	the interval about the measurement within which the true value is "confidently" expected to lie
coverage factor	-	the ratio of the expanded uncertainty to the combined standard uncertainty

The basic idea of Type A and Type B uncertainty (not to be confused with Type I and Type II errors of statistical inference) is to quantify "random" and "systematic" errors as standard deviations. The expanded uncertainty is nothing but the square root of the sum of the variances, and the coverage factor is nothing more than the number of standard uncertainties in the interval encompassing the measurement. By "international convention" NIST adopts a coverage factor of 2. If the reader has suffered through this chain of definition, he now realizes that the NIST policy boils down to: (1) e in equation (4-1) is a standard deviation; (2) e is made up of random and systematic components, each of which is quantified by its own standard deviation. None of the philosphical issues raised above is addressed, and there are theoretical reasons prohibiting expression of systematic error as a standard deviation. There appears to be no wholesale rush in water-quality monitoring to embrace the NIST policy.

The uncertainty may vary with the magnitude of the measurement, and the dependency may be generalized as

$$e = a + mv \qquad \text{for } v_0 < v < v_1 \qquad (4\text{-}2)$$

Actually, e may vary nonlinearly with mean value v of the measured parameter, but for present purposes an at-most-linear variation is sufficient (because the limited data usually available on precision will not support the assignment of a nonlinear variation). This formulation calls explicit attention to a range of applicability of the measurement from v_0 to v_1. Any analytical method has limits on its range of validity, though for some procedures these limits are so broad relative to the natural range of the variate that they are non-limiting in practice.

For a specific parameter, often the constant term a or the linear variation m v will dominate the dependency of error e on variate value v, and the other can be neglected. In the case of the former, the precision is constant over the range of applicability, and may be expressed simply as a constant value with the units of v. In the case of the latter, e may be conveniently stated as a fraction (a percentage) of v. (The suggestion appearing in recent editions of *Standard Methods* to report standard deviation as a percentage of the mean is unfortunate, in that it suggests that e varies directly with v , when in fact it may not.) Frequently there is inadequate data to determine which, if either, dominates. Sometimes, both may be important. For example, in Skougstad et al. (1979), the analysis for sediment boron is stated to have a precision (as a standard deviation) of about 7 mg/kg at the lower end of the range of applicability at 10 mg/kg, and about 50 mg/kg at the upper end of the range at 250 mg/kg. Substitution of these values in the above equation yields: $a \approx 5$ mg/kg and $m \approx 20\%$, so that the total precision is

$$e = 5 + 0.20v \qquad \text{mg/kg}$$

It is of course even better if there are multiple values of e for a range of values of v, whereupon a regression line can be estimated, and the best-fit values of a and m determined statistically. This is the format used in the most recent USGS manual (Fishman and Friedman, 1989) for dissolved analyses (see also Friedman and Erdmann, 1982), and when data warrant in the ASTM Annual.

One element in the above formulation that has special significance is the threshold value v_0 , which v must exceed for the analysis to be meaningful. Even for measurements that apply down to a value of zero, such a threshold value always exists, due perhaps to mechanical friction in a gauge or the limits of resolution of a probe. It may be much smaller than the lowest value of v encountered, or be much smaller than e for $v = 0$, and thus be practically negligible. This threshold value is referred to as the *detection limit*, and its operational definition is the value of analyte concentration that can be discriminated from the value determined from a laboratory blank at some pre-set probability. Clearly, in order to determine this, one must know how the analytical method behaves statistically for blanks, and all of the procedures for determining detection limits require data on statistics of measurements of blank samples. It has long been traditional in analytical chemistry (Vandecasteele and Block, 1993) to use a probability level of 99% that a value exceeding the detection limit represents a non-zero concentration, and this is in fact the definition adopted by EPA (EPA, 1992, Kimbrough and Wakakuwa, 1993) for determination of Method Detection Limits for its various analytical procedures. Therefore, if s is the standard deviation for blanks, the so-called background standard deviation, i.e. $s = a$ in eqn. (4-1), then the detection limit is approximately $v_0 = 3$ s.

For trace concentration determinations based upon instrumental methods however, especially those of hazardous or toxic contaminants, the method detection limit of the analysis takes on a singular importance in the reporting procedure because of the practice of *censoring*. By definition a measured concentration exceeding the level of 3s has a 1% chance of being in fact blank (zero). For measured concentrations less than 3s, the probability rises that these could in fact be zero. The practice is to report these values as "nondetects," i.e. below detection limits (BDL). The analytical procedure, it should be noted, will produce numbers less than the detection limit, even negative numbers. They are simply not reported quantitatively. Censored values present a great problem in water-quality analysis, detailed in Section 4.3.2 below, and their use has been criticized (e.g., Porter et al., 1988, D'Elia et al., 1989).

We note in passing that if one is intent upon censoring data, then setting a detection limit to limit the risk of a "false positive" is only a part of the censoring problem. One should also address the companion question of the risk of "false negatives," i.e. reporting as a nondetect when the analyte is in fact nonzero. For example, if the real concentration is exactly equal to the method detection limit (as defined by EPA, above), the measured concentrations for repeated measurements will scatter about this value, roughly half being less than the real value. Therefore, half of the measurements will be reported as nondetects. There is, therefore, both a type-I and a type-II delineation that must be made, and critical probability levels assigned to each. How this is handled has led to a number of alternative definitions of "critical level," "decision limit,""detection limit," "level of quantitation," "determination limit" and others (e.g., Vandecasteele and Block, 1993).

In order to completely characterize a measurement, we must include an estimate of the uncertainty, including any limiting values, such as the detection limit. In the Galveston Bay Status and Trends study (Ward and Armstrong, 1992a), a considerable effort was invested in determining the uncertainty of each of the analytes. The present project did not have the resources to repeat the effort of Ward and Armstrong (1992a), nor was it considered necessary since many of the data sources and parameters were the same in both studies.

Determination of parameter uncertainties was approached by Ward and Armstrong (1992a) in several ways depending upon the extent of documentation for the data set, in decreasing order of preference:

(a) review of QA/QC procedures observed by the collecting agency, as reflected in practices memos, manuals and directives,

(b) identification of the specific methodologies used and their established accuracy,

(c) statistical variation of the measurements themselves, relative to some external standard, e.g. a more accurate proxy relation or data from a contemporary, independent source.

(d) judgement, based upon experience with the method or equipment, and upon the practice of workers in the field using that methodology, as inferred from their explicit or implicit uncertainty statements.

For recent data with well-established procedures and QA/QC protocols, this was generally straightforward (though many agencies have no written descriptions and information had to be obtained from personal communications). For older data, the methodologies and probable care of the observers had to be judged. Published sources of precision data for specific analytical methods were used, especially *Standard Methods*, the ASTM annuals, and the USGS Techniques of Water-Resources Investigations. Generally, there was more information—and more quantitative scope—on precision in the later editions than the earlier, which raised a dilemma: when precision information changed, should the data contemporaneous with the measurements be used, i.e. assumed to be reflective of the technology and procedures of the time, or the more recent data derived from a larger base of measurements and presumably representing an improved estimate of precision applicable to the older techniques as well? Considering that the reported precision for many trace metals and organics is *lower* (i.e., greater standard deviations) in more recent publications (e.g., Fishman and Friedman, 1989) than in the older (e.g., Skougstad et al., 1979), this is not a merely pedantic concern. No doubt there are elements of truth in either alternative, but the former was elected. Ward and Armstrong (1992a) note that this is not an irreversible decision, as any later user of the data base has the option of employing a different measure of precision. The basic measurement itself is of course unaffected by the level of uncertainty assigned to it.

Also, we note that the precision data available is generally much more complete and accurate for the water-phase analytes than the sediment. Indeed, in the USGS manuals (Wershaw et al, 1987, Fishman and Friedman, 1989), for each of the bottom-material analyses there is simply the statement: "It is estimated that the percent relative standard deviation for [parameter name] in bottom material will be greater than that reported for dissolved [parameter name]." When precision data are presented for water-suspended sediment mixtures, we have used that preferentially over the dissolved data to estimate uncertainty for the sediment analysis.

Tables A-1 and A-2 in the appendix summarize the measures of uncertainty assigned in this study. As noted above, these uncertainty criteria were drawn from the compilation of Ward and Armstrong (1992a) for the Galveston Bay National Estuary Program, based upon available information on precision of various methodologies and procedures for different parameters (summarized in Table A-6 of the Appendix to Ward and Armstrong, 1992a, but not repeated in this report), suitably rounded and supplemented by estimates of accuracy from analysis of data from Galveston Bay when available, by precision data from similar compounds if primary data were not available (e.g., total DDT estimated from precision data for p,p'-DDT), and judgement of the Principal Investigators when no solid information was available.

In this section we have concentrated upon the analytical uncertainty of a measurement as though that is the central control on the accuracy of the data. The comparison of alternative methods of measuring salinity given in Section 2.1 above should be considered. Each of the methods of chlorides titration and conductivity measurement should be capable of yielding analytical accuracy to better than ±0.1‰, yet the actual realized accuracy is much poorer than this. This reinforces two points. First, we rarely have the luxury of simultaneous determinations of two related variables, by which we can evaluate the consistency and probable error of the data, as is the case in Figs. 2-1 *et seq*. What then of the many programs in which only a single measure of salinity, say, was made, and there is no means of cross-checking the data? Second, the precision of the methodology notwithstanding, it is the procedures and technique of the field crew, the laboratory and the data-entry personnel that are controlling in the level of accuracy attained. Even for as straightforward and commonplace a measurement as reading a conductivity meter or titrating for chlorides, the potential for error is substantial, as shown in Section 2.1. What then can be expected of more complex and demanding analyses of trace metals or organics? Any data point should be regarded with suspicion, and the cross-comparison with other nearby, contemporaneous measurements, even from different programs, should be an indispensable guide to weighing the reality of a measurement.

4.2.4 Data rejection

A separate concern in data processing is the handling of anomalous values lying well beyond the expected range of the variate. Most of these are the result of human error at some point in the process from laboratory or field measurement to entry into the data base. A frequent manifestation is a decimal point mislocation, resulting in multiplying the true value by one or several orders-of-magnitude. A screening rule can be formulated to reject such points. The problem is how to assign a rejection trigger so as to exclude points certainly in error, but not to exclude points that happen to deviate widely from "normal" values, since such deviations may in fact be real and therefore significant.

It has become traditional in data processing to differentiate between values that are so extreme as to be rejected as "unlikely" (including "impossible") and those that are "unusual" but within the realm of possibility, see, e.g., Bewers et al. (1975). This is the approach recommended by Tetra Tech (1987) who provide "A" and "B" values for an extensive list of estuarine variables, corresponding respectively to "unusual" and "unlikely." It must be noted that the normal strategy is to use these limits to identify anomalous points *during the data analysis and entry process*, to provide feedback to the originators of the data for verification and correction. In the present study, there is no prospect of tracing back to the originator of the data (except for verifying data entry performed during this project), so we need to determine a criterion for data rejection. We also note that in the present study any such rejection trigger would be applied at the earliest to the compilation of the Derivative Data Files, see Section 4.3 below, not to the source data (except, of course, for the "patently obvious" category described earlier).

Generally, as a matter of personal philosophy, we reject very little data even in the formulation of the Derivative Files, and reserve further data screening for the specific analyses to which the Derivative Data Bases are subject. Data were rejected from the source files if the date or position were obviously impossible and there were no satisfactory means of judging the correct value. For compilation of the Derivative Data Bases, rather broad ranges of admissable values were defined as bounds, and any values outside these bounds were screened out. These bounds are given in Table 4-1, along with the number of measurements actually eliminated by their application. As is apparent from this table, generally we did not reject data at this stage based on the parameter value, but reserved that for later steps in the analysis. For most of the parameters, there is no rejection whatever. Where there is rejection, the bounds are liberal enough that only extremely high values are affected, which are assuredly incorrect. Some, like the temperature of 100°C or the DO of 69 mg/L, are impossible. Four sediment volatile solids exceed 20%. Though lying within the range of *possibility*, these lie outside the range of *probability*. (Surely there would have been reports of lab technicians lacking eyebrows.) Beyond these unambiguous cases, though, the demarcation between probable and improbable is less certain: the 10 ppb values for water-phase 2,4,5-T and Silvex, or the 640 μg/kg of sediment BaP may be more unlikely than unusual, but we are hesitant to dismiss them on strictly an *a priori* basis.

Rejection triggers for later stages in the analysis were assigned to many (not all) of the variables based upon the suggestions of Tetra Tech (1987) or on judgement of the PI's. These are given in Tables A-1 and A-2 of the Appendix. (We note that some of the values in the Tetra Tech report are inapplicable to Corpus Christi Bay, e.g., a temperature limit of 30°C.) Both the uncertainty and the rejection triggers are provided more as guidance to the future users of these data sets than as absolute bounds on data inclusion, and reflect as much our judgement of the quality of the different data programs as statistical constructs.

Data rejection can be performed based upon either the level of uncertainty of the measurement or its magnitude relative to the rejection trigger (when one is provided). Each measurement in the Derivative Data Base is accompanied by the specified level of confidence, transformed into units of the variable and scaled (when appropriate) to the magnitude of the measurement. Thereafter, any data processing can be preceded by an assignment of acceptable accuracy of measurement; any measurements failing this level would be excluded from that analysis. But these measurements would still be retained in the data base. We believe this to be a superior approach to merely deleting data, especially older data, by a sharply defined criterion of "reliability." This is closely related to the notion of preservation of data integrity discussed above.

A separate problem is presented by the presence of anomalous zero values for trace organics and metals, primarily in the TWDB and SMN data bases from the 1970's. This is discussed in 5.2.2 below. One source for these zeroes, we believe, is replacement of "censored" analytical reports, that the analyte was "below detection limits," with a value of zero. To retain the zero values would clearly bias trend analyses, since "detects" become more frequent later in the period of record due to improved analytical accuracy, which would create an increasing trend line with no basis in reality. Yet to delete these from the data base would be to lose potential information. While the zeroes were allowed to stand as reported in the Derivative Data Bases, for the analyses we replaced these with "below detection limit" flags. Unfortunately, no information has survived as to the actual detection limits of the methods used. Therefore, we postulated detection limits based upon the laboratory methodologies current at the time (see Appendix A-6 of Ward and Armstrong, 1992a), and/or reported by contempo-raneous labs in other data sets. The parameters affected and these estimated detection limits are tabulated in Table 4-2.

4.3 Data Set Processing

4.3.1 Preparation of data files

The principal steps in data-processing in this study to create data files suitable for statistical analyses were:

(1) For each historical data program in the Corpus Christi Bay system, compile a Primary Data File, consisting of the digital record of measurements ordered chronologically;

(2) For each parameter of concern, sift through the Primary Data Files, applying whatever screening, proxy relationships, and units conversions are necessary, to create a Master Derivative File for that parameter;

Table 4-1
Acceptable-range bounds for water/sediment parameters
for inclusion in Derivative Data Bases

parameter / units		*bounds*		*rejected**	*parameter / units*		*bounds*		*rejected**
WQALK	ppm	0	1000		WQTPO4	ppm	0	50	
WQAMMN	ppm	0	50		WQTRANS	1 m %	0	100	1
WQBOD5	ppm	0	100		WQTSS	ppm	0	5000	
WQCHLA	ppb	0	2000	2	WQTURBJ	JTU	0	1000	
WQCHLB	ppb	0	50		WQTURBN	NTU	0	1000	
WQCYAN	ppm	0	50		WQVOLS	ppm	0	10000	
WQDO	ppm	0	25	5	WQVSS	ppm	0	500	
WQFCOLI	MPN/200ml	0	500000		WQmetagd	ppb	0	100	
WQKJLN	ppm	0	50		WQmetagt	ppb	0	1000	
WQNO2N	ppm	0	5		WQmetasd	ppb	0	100	
WQNO3N	ppm	0	20		WQmetast	ppb	0	1000	
WQO&G	ppm	0	100		WQmetbt	ppb	0	5000	2
WQOPD	ppm	0	20		WQmetbd	ppb	0	5000	5
WQOPO4	ppm	0	50		WQmetbad	ppb	0	1000	4
WQORGN	ppm	0	10		WQmetbat	ppb	0	1000	
WQPH		4	12	3	WQmetcdd	ppb	0	50	
WQPHEO	ppb	0	1000		WQmetcdt	ppb	0	100	4
WQSAL	ppt	0	350		WQmetcod	ppb	0	10	
WQSECCHI	m	0	15		WQmetcot	ppb	0	50	
WQSIO2	ppm	0	50		WQmetcrd	ppb	0	100	
WQSO4	ppm	0	8000	5	WQmetcrt	ppb	0	500	
WQTCOLI	MPN/200ml	0	500000		WQmetcud	ppb	0	100	
WQTEMP	deg C	0	45	4	WQmetcut	ppb	0	500	
WQTOC	ppm	0	100	17	WQmetfed	ppb	0	500	
WQTOTP	ppm	0	50		WQmetfet	ppb	0	10000	

(continued)

Table 4-1
(continued)

parameter / units		*bounds*		*rejected**	*parameter / units*		*bounds*		*rejected**
WQmethgd	ppb	0	5		WQ-CHLR	ppb	0	10	
WQmethgt	ppb	0	20		WQ-CHLRC	ppb	0	10	
WQmetmnd	ppb	0	500		WQ-CHRYS	ppb	0	10	
WQmetmnt	ppb	0	4000		WQ-DBANE	ppb	0	10	
WQmetnid	ppb	0	50		WQ-DDD	ppb	0	10	
WQmetnit	ppb	0	500		WQ-DDE	ppb	0	10	
WQmetpbd	ppb	0	50		WQ-DDT	ppb	0	10	
WQmetpbt	ppb	0	500		WQ-DIAZ	ppb	0	10	
WQmetsed	ppb	0	5		WQ-DIEL	ppb	0	10	
WQmetset	ppb	0	50		WQ-ENDO	ppb	0	10	
WQmetsrd	ppb	0	10000	4	WQ-ENDR	ppb	0	10	
WQmetznd	ppb	0	1000		WQ-FLRA	ppb	0	10	
WQmetznt	ppb	0	5000		WQ-FLRN	ppb	0	10	
WQ-245T	ppb	0	10		WQ-HEPT	ppb	0	10	
WQ-24D	ppb	0	10	1	WQ-HEPX	ppb	0	10	
WQ-ABHC	ppb	0	10		WQ-HEXA	ppb	0	10	
WQ-ACEN	ppb	0	10		WQ-I123P	ppb	0	10	
WQ-ACENA	ppb	0	10		WQ-LIND	ppb	0	10	
WQ-ALDR	ppb	0	10		WQ-MALA	ppb	0	10	
WQ-ANTHR	ppb	0	10		WQ-MTHP	ppb	0	10	
WQ-BNZA	ppb	0	10		WQ-MTHX	ppb	0	10	
WQ-BNZAA	ppb	0	10		WQ-NAPT	ppb	0	10	
WQ-BNZB	ppb	0	10		WQ-ODDT	ppb	0	10	
WQ-BNZGP	ppb	0	10		WQ-PAH	ppb	0	10	
WQ-BNZK	ppb	0	10		WQ-PARA	ppb	0	10	

(continued)

Table 4-1
(continued)

parameter / units		*bounds*		*rejected**	*parameter / units*		*bounds*		*rejected**
WQ-PCB	ppb	0	10		sedmetcu	ppm	0	300	
WQ-PCP	ppb	0	10		sedmetfe	ppm	0	50000	
WQ-PDDD	ppb	0	10		sedmethg	ppm	0	50	
WQ-PDDE	ppb	0	10		sedmetmn	ppm	0	1500	
WQ-PDDT	ppb	0	10		sedmetni	ppm	0	300	
WQ-PHNAN	ppb	0	10		sedmetpb	ppm	0	300	
WQ-PYRN	ppb	0	10		sedmetse	ppm	0	50	
WQ-SLVX	ppb	0	10		sedmetsr	ppm	0	7000	
WQ-TOXA	ppb	0	10		sedmetzn	ppm	0	10000	1
sedcyan	ppm	0	2		sed-245t	ppb	0	50	
sedkjln	ppm	0	10000	1	sed-24d	ppb	0	50	
sedo&g	ppm	0	10000	3	sed-abhc	ppb	0	50	
sedammn	ppm	0	1500		sed-acen	ppb	0	1000	
sedorgn	ppm	0	1500		sed-acyn	ppb	0	1000	
sedtoc	ppt	0	100		sed-aldr	ppb	0	50	
SEDtotp	ppm	0	10000		sed-anth	ppb	0	1000	
sedvols	ppm	0	200000	4	sed-bnza	ppb	0	1000	
sedmetag	ppm	0	100		sed-bnze	ppb	0	1000	
sedmetal	ppm	0	50000		SD-bnzaa	ppb	0	1000	
sedmetas	ppm	0	100		SD-bnzb	ppb	0	1000	
sedmetb	ppm	0	500		SD-bnzk	ppb	0	1000	
sedmetba	ppm	0	3000	3	SD-bnzgp	ppb	0	1000	
sedmetcd	ppm	0	300		sed-chlr	ppb	0	100	
sedmetco	ppm	0	50		sd-chlrc	ppb	0	50	
sedmetcr	ppm	0	300		sed-chry	ppb	0	1000	

(continued)

Table 4-1
(continued)

parameter / units		*bounds*		*rejected**	*parameter / units*		*bounds*		*rejected**
WQ-PCB	ppb	0	10		sedmetcu	ppm	0	300	
WQ-PCP	ppb	0	10		sedmetfe	ppm	0	50000	
WQ-PDDD	ppb	0	10		sedmethg	ppm	0	50	
WQ-PDDE	ppb	0	10		sedmetmn	ppm	0	1500	
WQ-PDDT	ppb	0	10		sedmetni	ppm	0	300	
WQ-PHNAN	ppb	0	10		sedmetpb	ppm	0	300	
WQ-PYRN	ppb	0	10		sedmetse	ppm	0	50	
WQ-SLVX	ppb	0	10		sedmetsr	ppm	0	7000	
WQ-TOXA	ppb	0	10		sedmetzn	ppm	0	10000	1
sedcyan	ppm	0	2		sed-245t	ppb	0	50	
sedkjln	ppm	0	10000	1	sed-24d	ppb	0	50	
sedo&g	ppm	0	10000	3	sed-abhc	ppb	0	50	
sedammn	ppm	0	1500		sed-acen	ppb	0	1000	
sedorgn	ppm	0	1500		sed-acyn	ppb	0	1000	
sedtoc	ppt	0	100		sed-aldr	ppb	0	50	
SEDtotp	ppm	0	10000		sed-anth	ppb	0	1000	
sedvols	ppm	0	200000	4	sed-bnza	ppb	0	1000	
sedmetag	ppm	0	100		sed-bnze	ppb	0	1000	
sedmetal	ppm	0	50000		SD-bnzaa	ppb	0	1000	
sedmetas	ppm	0	100		SD-bnzb	ppb	0	1000	
sedmetb	ppm	0	500		SD-bnzk	ppb	0	1000	
sedmetba	ppm	0	3000	3	SD-bnzgp	ppb	0	1000	
sedmetcd	ppm	0	300		sed-chlr	ppb	0	100	
sedmetco	ppm	0	50		sd-chlrc	ppb	0	50	
sedmetcr	ppm	0	300		sed-chry	ppb	0	1000	

(continued)

Table 4-1
(continued)

parameter / units		*bounds*		*rejected**	*parameter / units*		*bounds*		*rejected**
sed-ddd	ppb	0	100		sed-napt	ppb	0	1000	
sed-dde	ppb	0	50		sed-pah	ppb	0	5000	
sed-ddt	ppb	0	50		sed-para	ppb	0	50	
sed-diaz	ppb	0	50		sed-pcb	ppb	0	50	
SD-dbane	ppb	0	1000		sed-pcp	ppb	0	50	
sed-diel	ppb	0	50		sed-pddd	ppb	0	50	
sed-endo	ppb	0	50		sed-pdde	ppb	0	50	
sed-endr	ppb	0	50		sed-pddt	ppb	0	50	
sed-flra	ppb	0	1000		sed-oddt	ppb	0	50	
SD-flrn	ppb	0	1000		sed-oddd	ppb	0	50	
sed-hept	ppb	0	50		sed-odde	ppb	0	50	
sed-hepx	ppb	0	50		sed-pery	ppb	0	1000	
sed-hexa	ppb	0	50		SD-phnan	ppb	0	1000	
SD-I123p	ppb	0	50		SD-pyrn	ppb	0	1000	
sed-lind	ppb	0	50		SED-slvx	ppb	0	50	
sed-mala	ppb	0	50		sed-toxa	ppb	0	50	
sed-mthp	ppb	0	50		sed-tbt	ppb	0	500	

* Blank signifies no measurements rejected

Table 4-2
Estimated detection limits for parameters with zero values in CCBNEP data base
(See Table 2-1 for parameter name definitions and measurement units)

parameter	*DL*	*parameter*	*DL*
WQmetagd	1	WQ-MTHP	0.2
WQmetasd	1	WQ-MTHX	1
WQmetcdd	1	WQ-PARA	0.5
WQmetcdt	2	WQ-PCB	0.1
WQmetcod	1	WQ-SLVX	0.01
WQmetcot	2	WQ-TOXA	0.5
WQmetcrd	1	sedmetcd	0.1
WQmetfed	20	sed-245t	20
WQmetfet	20	sed-24d	0.5
WQmethgd	0.1	sed-aldr	0.2
WQmetmnd	10	sed-chlr	1
WQmetnid	1	sd-chlrc	20
WQmetpbd	1	sed-ddd	0.2
WQmetpbt	2	sed-dde	0.2
WQmetsed	1	sed-ddt	0.2
WQmetznd	2	sed-diaz	5
WQ-245T	10	sed-diel	0.2
WQ-24D	50	sed-endr	0.2
WQ-ALDR	0.02	sed-hept	1
WQ-CHLR	0.1	sed-hepx	1
WQ-DDD	0.01	sed-lind	1
WQ-DDE	0.01	sed-mala	5
WQ-DDT	0.01	sed-mthp	5
WQ-DIAZ	0.01	sed-mthx	20
WQ-DIEL	0.02	sed-para	5
WQ-ENDO	0.01	sed-pcb	2
WQ-ENDR	0.01	sed-pddt	5
WQ-HEPT	0.01	sed-pery	3
WQ-HEPX	0.06	SED-slvx	20
WQ-LIND	0.03	sed-toxa	50
WQ-MALA	1		

(3) Sort the Master Derivative Files into geographic segments for the Corpus Christi Bay system;

After Step 3, for each parameter there would be created a chronological record of measurements of that parameter for each geographical segment of the bay. This is now considered to be an autonomous data set which can be subjected to various additional pairing, sorting and statistical analyses as necessary to expose time-space variations. The general processing procedures are shown schematically in the flow charts of Figs. 4-1 through 4-3.

The construction of the Primary Data File was described in Section 4.1 above. Because the ultimate product is to be a chronological Derivative File, and the process of chronologizing a file can be resource-intensive, it was decided to chronologize the data records as early in the process as feasible, then to design subsequent data handling in such a way that ordering is preserved. This is the principal difference between the Primary File and the Source File, the digital record in the format and units of the agency that obtained the data, i.e. the Primary File contains exactly the same data records except ordered chronologically. A secondary difference is that the first tier of the Q/A screening is applied in this process, see Fig. 4-1, so that the Primary File will have entry errors and "patently obvious" data errors corrected or deleted.

The creation of the Derivative Data Files is fundamentally a matter of merging information from various files and re-formatting the product. The various steps in this procedure are shown in Fig. 4-2. The sampling station latitude/longitude coordinates are collected in a separate file, and accessed according to the agency station designations to merge the coordinates with the data taken at that station. At this stage, all units conversions are applied, as well as any proxy relationships by which one parameter may be transformed into another. Because we anticipate analyzing data on a time scale of days to weeks, the information on clock time (i.e., time of day) of each sample is not carried through to the Derivative Data Files, but the full date is retained. In addition to the parameter value itself, the uncertainty is estimated and included in the data record.

The format of each record in the Derivative Data Files is:

DATE LATITUDE LONGITUDE DEPTH MEASRMT UNCRNTY PRJ

where DATE, LATITUDE AND LONGITUDE are 6-digit fields (YRMODA and the latitude/longitude coordinates are degrees/minutes/seconds), the sample depth is in meters, MEASRMT is the measured value of the parameter (retaining three significant figures), UNCRNTY is the uncertainty as a standard deviation following the convention of Section 4.2.3 above (to two significant digits), and PRJ is a 3-digit integer flag that identifies the agency or project that was the source of the measurement. Thus, each record of the Derivative Data File represents a point in time (to resolution of a day) and space (horizontal and vertical position), together with the measurement and its uncertainty. Each such record requires 50 bytes of storage.

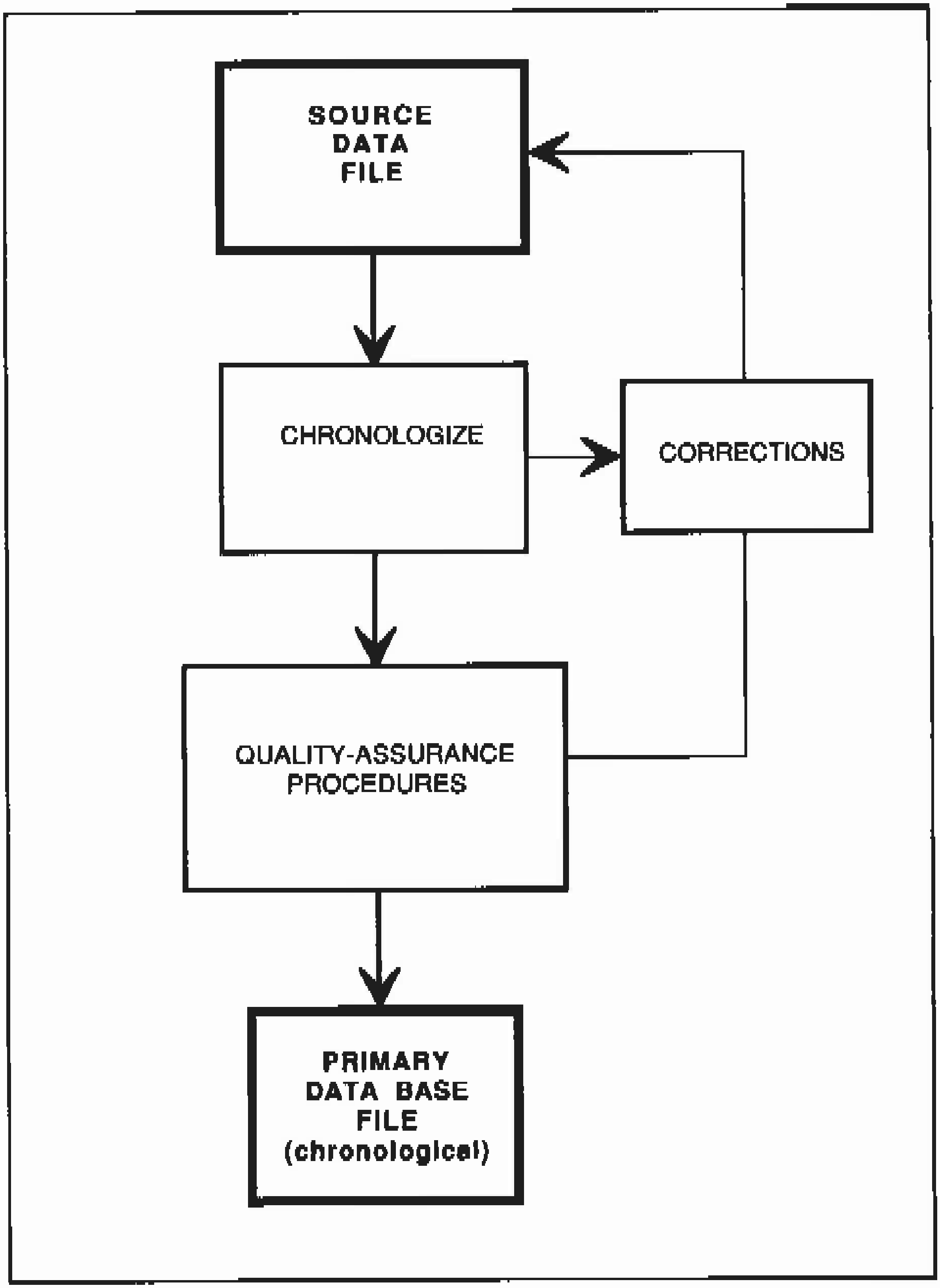


Fig. 4-1 Procedure for pre-processing and generation of primary data file

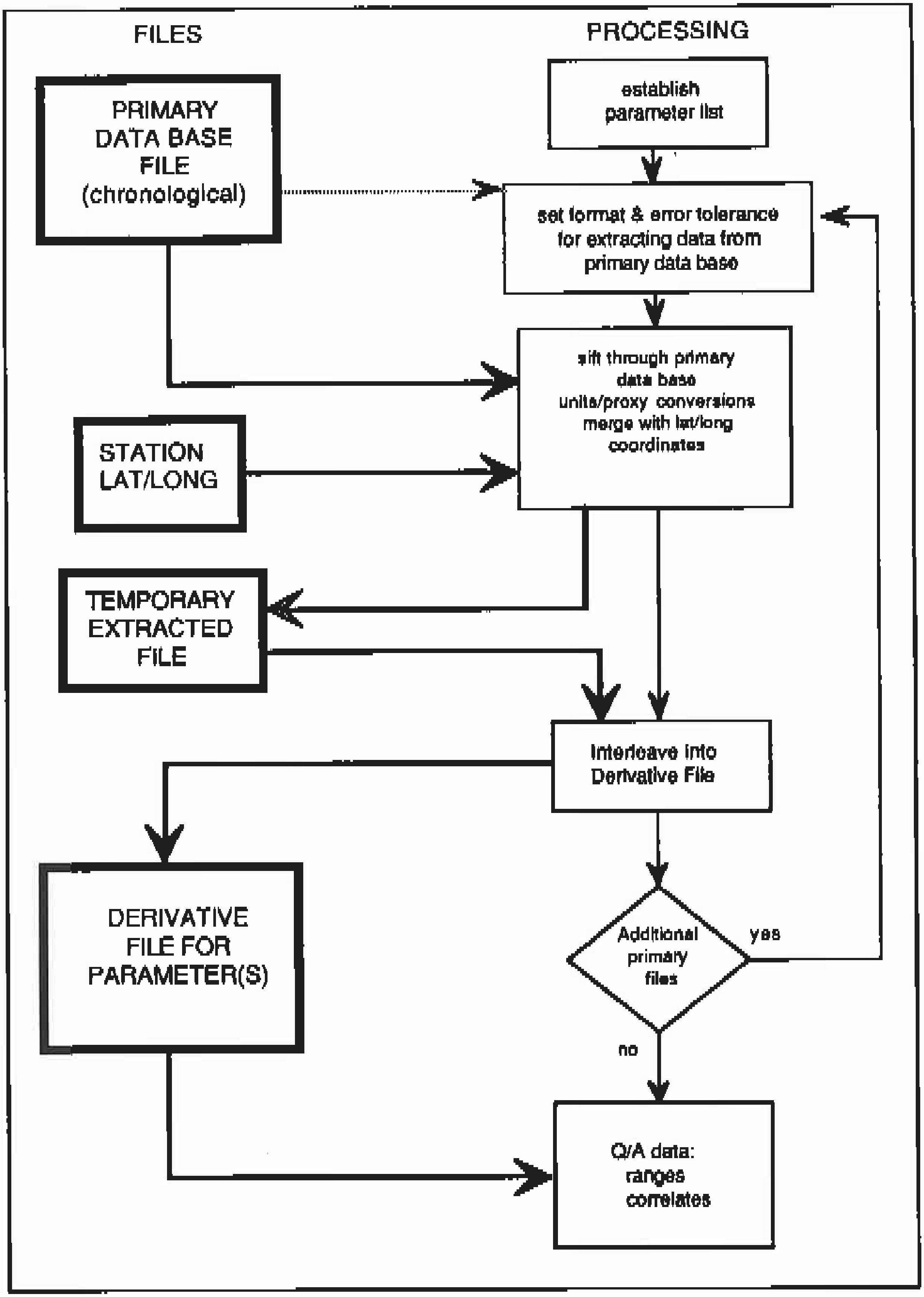


Fig. 4-2. Procedure for creation of derivative data bases

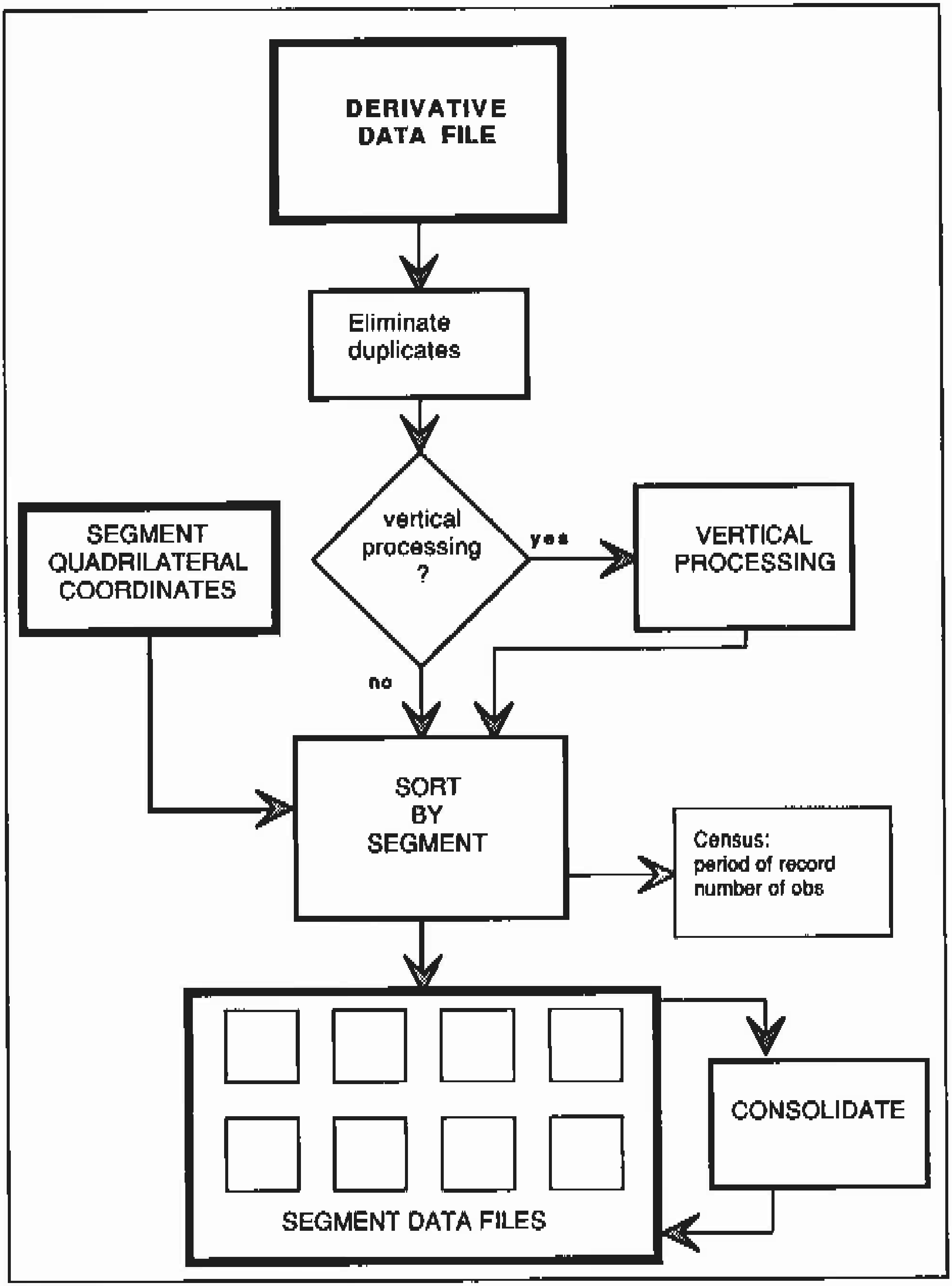


Fig. 4-3. Post-processing of Derivative Data File

Throughout this process there are numerous error traps and cross-checks, not only to ensure that the data is not corrupted by a bug in the processing but also to detect entry errors or aberrancies in the data as reported by the agency. The temporary extracted file shown in Fig. 4-2 contains the data in the above format from a single source. This file is examined closely for errors or anomalies before it is interleaved into the Master Derivative File. Once the Derivative File is created, it can be subjected to various screenings and data rejection, according to the preferences of the researcher.

We regard the Master Derivative Data files to be our principal data resource product from this study. These contain all of the data for each parameter that we were able to locate and digitize, and incorporate our judgment on which data should be retained or rejected. In order to address the concerns of characterizing the ambient quality of Corpus Christi Bay and its historical trends, these data bases are subjected to additional processing, as indicated in Fig. 4-3. Principally, this involves further filtering of the data and sorting the data into the separate geographical segments. Because the source data files can contain duplicate measurements, e.g. the TWDB Coastal Data System may contain TWC measurements that are also in the Statewide Monitoring Network system, some research projects may share the same data files, keyboarding personnel may have inadvertently duplicated entries, etc., there is the possibility that duplicate measurements may be present in the file. Therefore, there is a preliminary screening step to detect such duplicates. (This is repeated after the data are sorted by segment to detect "near-duplicates.") "Vertical processing" in Fig. 4-3 refers to selection of data from only one depth (or range of depths) or to averaging data in the vertical.

Finally, the defining quadrilaterals for a system of segmentation are applied to the data to sort into the various segments. For this project, two segmentation systems were employed, as discussed in Chapter 3. The first is the Texas Natural Resource Conservation Commission water quality segments, which are defined and their associated geographical quadrilaterals given in Tables A-3 and A-4, resp., in the Appendix. The second is the system of hydrographic segments shown in Figs. 3-3 through 3-7 and defined by the quadrilaterals of Table A-5 in the Appendix. As some of the TNRCC segments are represented by the union of two or more quadrilaterals, once the initial sorting is completed these quadrilaterals must be consolidated into a single data record for that segment, hence the process of "consolidation" in Fig. 4-3. At this point, the data files are in a form suitable for statistical analysis.

4.3.2 Data analyses

As described above, once the data is sorted into segments, for each parameter and each segment, there is created a data record for that segment, considered as an autonomous data file. These data may have been further screened in the process, e.g. by limiting only to "near-surface" values. Also various derivative calculations may have been carried out to create new data files, for example vertical stratification or dissolved oxygen deficit. The next stage in the processing is the analysis of these individual data files.

Generally, a two-tier statistical analysis was performed. The first is descriptive in nature, computing statistical indicators such as arithmetic average, standard deviation and variance, and extrema. The second is to determine historical trends by methods of linear regression. All of these are discussed in more detail in Chapter 6 in association with the interpretation of the results. The results of all of these analyses are presented in tabular form in Appendices B and C, for both the TNRCC segmentation and the Hydrographic Area segmentation, respectively.

It is worth-while to re-emphasize the distinctions between these segmentations. The TNRCC segmentation carries the weight of tradition, having formed the basis for water quality management in the system for many years. The segments tend to be rather coarse in space, however, as evidenced by the fact that the study area, including several freshwater tributaries, total 43 segments, counting 10 that were added to the 32 specified in the original project scope of work. The hydrographic areas, on the other hand, are much more spatially refined, totalling 160 in the estuarine portion of the study area, and 18 in the nearshore Gulf of Mexico. However, these generally do not include the freshwater tributaries or the tidal reach of these tributaries.

The treatment of detection limits in analysis of water quality is particularly vexing. As noted in Section 4.2.3 above, the practice in analytical chemistry has been to censor data whose magnitude is small enough that the probability of it differing from zero exceeds some preset threshold, typically 1%. This is the detection limit, and the measurement is replaced by the statement that the value is below detection limits. In the data collected roughly since the mid-1970's, detection limits are generally reported as part of the data set. Thus a real measurement (with, of course, an associated uncertainty) is replaced with a "statement." The analytical problem is how to recover quantitative information from the simple statement that the concentration is "below detection limits."

There are three logical alternatives, each of which has a rational basis. First, the measurements BDL can be simply ignored, as providing essentially no quantitative information. Second, the BDL values can be replaced with zero in the analyses, on the argument that for practical purposes the parameter is not present. This is probably the most commonly elected alternative. It is, for example, the approach adopted by the National Ocean Service in its National Status & Trends Program (NOS, 1991). Third, the BDL values can be taken to be the reported detection limits, on the basis that the actual concentration could be as high as the detection limit.

In our view, the selection is dependent upon the purpose at hand. The non-BDL statistics can provide some insight into the precision and variability of the parameter, which the more constant DL values would corrupt or even mask. However, to completely ignore BDL results is to lose information, albeit non-quantitative. The fact is that a water or sediment sample was obtained (usually at great effort), a careful analysis performed, and an upper bound established on the concentration of the parameter. This information should not be dismissed cavalierly. The latter two alternatives use that information, either optimistically or pessimistically, depending upon the intent of the analyst. In this project, with typical equivocation, we decided to employ all three, i.e. to compute *appropriate* statistics with only above-DL data, with the BDL values set to zero and with the BDL values set to the DL, thereby establishing a probable *range* of the statistic. The "appropriate" statistics include averages and variability for the above-DL data, but do not include calculations of variability for the latter two, since the largely invariant values of either end of the the range (i.e. either value assumed for a BDL measurement) would distort the results. Even in a trends analysis (which is variability in time), to incorporate 0 or DL values might either mask any vestige of a real trend by padding the data with zeroes or displace the real trend with a trend of measurement sensitivity. The user of these results therefore can choose among them whichever best serves the purpose of the analysis.

We note in passing that other options for treating BDL's exist in the literature. One is to replace the BDL with a value that is equal to one-half the detection limit. If one doesn't believe in sea monsters, would one accept half a sea monster? On a more sophisticated level, there are theories in which the values above detection limits are used to fit the lower tail of a probability distribution, usually log-normal, which is then sampled to incorporate the measurements below detection limits. Would one accept the existence of sea monsters if extrapolated from the existence of fish? These theories (see, e.g., Gilbert, 1987, Gilliom and Helsel, 1984, Helsel, 1990) are an outgrowth of probit theory (Finney, 1952), and clearly work better when the BDL's are in the minority of the measurements. In any event, the two assumptions employed in the present study (setting BDL's =0 and setting BDL's = detection limit) will bound the answer that would have been obtained from these more complicated analyses. For present purposes, we believe that is sufficient.

5. THE DATA BASE FOR CORPUS CHRISTI BAY

5.1 Data sources and data acquisition

The present Status & Trends project was in many ways at a disadvantage in comparison to its earlier counterpart in the Galveston Bay National Estuary Program (GBNEP). The study area is much larger, the scope of the program was enlarged, both in terms of target parameters and in including tissue data as an additional scope-of-work item, and the budget and study period were both reduced. Moreover, the GBNEP preceded its Status & Trends Project (Ward and Armstrong, 1992a) with a comprehensive one-year data-inventory project (Ward and Armstrong, 1991) during which the major sources for data were identified and inventoried, and some of the rare, inaccessible data sets were acquired. But one significant advantage to this project deriving from the earlier work in Galveston Bay is that many of the data sources and contacts are the same for the Corpus Christi Bay system, including the Texas Natural Resource Conservation Commission (TNRCC), Texas Water Development Board (TWDB), Texas Parks and Wildlife Department (TPWD), Texas Department of Health (TDH) and U.S. Corps of Engineers (USCE). Just as important, the GBNEP Data Inventory project entailed an enormous amount of agency and individual contacts, many of which proved to be unfruitful for data relating to water and sediment quality. In this project, therefore, the needless expenditure of energy in seeking data from these sources was avoided. Thus, for many of the key data sources for Corpus Christi Bay, we could proceed to acquire the data for Corpus Christi Bay much more efficiently, as beneficiaries of the GBNEP experience.

Another significant benefit to this project which we believe to have derived, at least in part, from the earlier Galveston Bay work is a heightened awareness of the lamentable state of preservation and management of older data among many of the resource agencies and workers. In some cases, this heightened awareness has led to improved data management, and/or a greater willingness to provide data to the present study. This is addressed in more detail in the concluding section of this chapter.

In the GBNEP Data Inventory work, an index of "information content" was proffered to provide a bases for quantifying the data resource as a function of time and discipline. This lay beyond the scope of the present study, but it is our intuition from working with the CCBNEP data sets that the same general conclusion would follow, that the data resource for the bay as a whole is dominated by a few large-scale collection activities, with numerous much smaller projects. This does not imply, however, that the smaller projects may be ignored. The *cumulative* information in these smaller studies can equal that in many larger projects. Further, these smaller projects may fill important gaps in the space-time record. Unfortunately, in a project of limited resources such as this one, the effort necessary to track down the raw data has to be tempered by the law of diminishing returns.

An important feature of prosecuting a project of this nature is the fact that most of the principal tasks in this project are *serial* in nature, i.e. data acquisition must precede data compilation, which in turn must precede statistical analysis, which in turn precedes cause-and-effect analysis. (There is some opportunity for parallel efforts during the various work tasks, but this is minor compared to the overall serial nature of the effort.) The central constraining milestone in the project schedule is the point of completion of the compiled digital data base. This is the schedule *hinge point*. The analytical phases of the project cannot begin until this point. After this point, no additional data can be added to the data base (without, of course, entailing a complete repetition of all the analytical tasks carried out thus far). The project was originally given a 12-month schedule (compared to 2.5 years required for the same work in the Galveston Bay project), and work on the project started several months in advance of the actual contract period. Despite this, the infeasibility of meeting the original schedule was almost immediately recognized. As literature review, agency contacts, data acquisition and data compilation proceeded, we continued to revise the schedule, and move the hinge point ahead, like a carrot receding before a hungry donkey.

One factor affecting prosecution of the work which unfortunately did not change from the experience in Galveston Bay was the generally unsatisfactory response to the inquiries of the Principal Investigators, especially among some regional agencies and individual researchers. Despite follow-up letters, telephone calls and e-mail, some of these responded only months after the original inquiry and some not at all. In fact a substantial amount of historical data, of immense potential value in extending the data density and period of record, are still in the hands of individual researchers, but we could not afford to delay this work any longer awaiting its receipt. Even after the data compilation was declared finished, and the work proceeded into the analysis phase, data sets continued to dribble in. Some of these, unfortunately, simply could not be included in the present analysis (though they will be incorporated into the master data base for future studies).

5.2 Data collection in Corpus Christi Bay

The data analyzed in this project were drawn from numerous past programs in Corpus Christi Bay. These programs are summarized in Table 5-1. Each of these comprises measurement of some of the water or sediment quality variables within a part of Corpus Christi Bay for some definite sampling interval and period. Apart from this general statement, the programs differ in objectives and procedures.

At the most basic level, we differentiate between the objectives of monitoring, survey and focused research. *Monitoring* programs are put in place for a protracted or indefinite period for the purposes of sampling a suite of variables. The data from such a program generally serves more than one purpose. A key characteristic of a monitoring program is consistency in the suite of variables acquired, since it is the accumulation of a long-period data base that is the

Table 5-1
Current and historical sampling programs in Corpus Christi Bay study area providing data for CCBNEP Status and Trends analysis

Code	*Abbre-viation*	*Agency or source*	*Project or Program*	*Format of source*	*Comments*
1	SMN	TNRCC	Statewide Monitoring Network	ASCII-coded data from TNRCC	Transferred by ftp via Internet, reformatted with special-purpose programs
2	CDS	TWDB	Coastal Data System	ASCII files	Multiple files, some combined some separated by contractor.
3	TPWD	TPWD	Coastal Fisheries Hydrographic obs	ASCII	Transported by diskette. Re-formatted with special-purpose codes
4	TGFOC	TPWD	Older hydrographic data from 1950's-1960's	hardcopy tables or field notes	Keyboarded by this project.
5	SWRI	Southwest Research Inst.	Corpus Christi Bay Project mid-1970's	hardcopy typed images of field sheets	Keyboarded by this project.
6	TDH EST	Texas Dept of Health	Estuarine Data File	ASCII zipped from TSDH	Re-formatted with special-purpose codes
7	TSDH	Texas Dept of Health	Water chemistry program of 1960's and 1970's	ASCII archive files	Re-formatted with special-purpose codes

(continued)

Table 5-1 (continued)

Code	*Abbreviation*	*Agency or source*	*Project or Program*	*Format of source*	*Comments*
8	USCE7	Corps of Engineers Galveston District	O&M Division water & sediment 1970s data	hard-copy tabulations	Keyboarded by this project.
9	USCE8	Corps of Engineers Galveston District	O&M Division water & sediment 1980s data	hard-copy, some LOTUS	Keyboarded by this project. Some re-formatted.
10	USCE9	Corps of Engineers Galveston District	O&M Division water & sediment 1990s data	delimited-text ASCII from USCE	Some keyboarded by this project.
11	USGS	U.S. Geological Survey Corpus Office	Sediment chemistry & some hydrographic data	Digital QUATROPRO	Major re-formatting with special-purpose codes
12	BEG	Bureau of Economic Geology, UT	Submerged Lands Study sponsored by GLO	ASCII files	Minor re-formatting required.
13	SWRI TRL	Southwest Research Institute	Hydrographic data from 1976-77 Trawl Study	hard copy tables	Keyboarded by this project.
14	TAMU -40	Dept. Oceanogr. Texas A&M Coll.	Reynolds Metals Baseline Study	hard copy tables	Keyboarded by this project.
15	OXY CHEM	CCB Foundation/ Oxychem	La Quinta Channel Survey	hard copy	Keyboarded by this project.

(continued)

Table 5-1 (continued)

Code	Abbre-viation	Agency or source	Project or Program	Format of source	Comments
16	MSI-LM	Marine Science Inst.—Whitledge	Laguna Madre nutrients data	digital EXCEL	Minor reformatting required.
17	MSI-NB	Marine Science Inst.—Whitledge	Corpus Christi & Aransas nutrients data	digital EXCEL	Minor reformatting required.
18	NOS	National Ocean Service of NOAA	National Status & Trends Project	digital download, but a mess	Extensive re-formatting required
19	USFWS	U.S. Fish & Wildlife Service	Corpus Christi Bay Project (late 1970's)	hard copy & graphical, very disorganized	Keyboarded by this project.
20	JMA	James Miertschin Associates	Hydrographic surveys of Nueces Bay 1990's	digital EXCEL	Minor reformatting required.
21	EMAP	Environmental Protection Agency	EMAP/REMAP	digital diskettes	Minor reformatting required.
22	CBI	Blucher Inst./ TAMU–CC	Hydrosonde data from CCBNEP area	digital ASCII Internet ftp	See discussion in text
23	TWDB	Texas Water Development Board	Hydrosonde data	digital ASCII Internet ftp	See discussion in text
24	MISC	TGFOC, Humble Oil, Sun Oil and others	Data from Laguna Madre surveys of 1940's and older data	hard copy	Keyboarded by this project. See discussion in text.

(continued)

Table 1 (continued)

Code	*Abbreviation*	*Agency or source*	*Project or Program*	*Format of source*	*Comments*
25	NCDH	Corpus Christi – Nueces County Health Department	Routine shoreline water quality surveys (coliforms only)	hard copy lab reports	Keyboarded by this project.
26	UTA-GEOL	Geology Dept Univ. of Texas	Masters thesis of Suter	hard copy tables	Keyboarded by this project.
27	USFWS-CCB	U.S. Fish & Wildlife Service, Corpus Christi lab	Contaminants assessment	Digital: spreadsheets	Minor reformatting required. (some parameters missing.)
28	TDH-TIS	Texas Department of Health	Seafood Safety Division tissue data	Hard copy tables	Keyboarded by this project.
29	TAMU-72	Texas A&M Univ. College Station	Estuarine systems Project	Hard copy tables in report	Keyboarded by this project.
30	TAMU-CCS	Texas A&M Univ. Corpus Christi Center for Coastal Studies	Nueces Marsh Mitigation Studies	Report and diskettes	Minor reformatting required.

purpose of the program. Important monitoring programs in Corpus Christi Bay include the Texas Natural Resource Conservation Commission Statewide Monitoring Network sampling, the Texas Parks and Wildlife Department Coastal Fisheries Program, and the Texas Department of Health Shellfish Sanitation data-collection program.

A *survey*, in contrast, is characterized by a definite limit in time. It may be a one-time sampling run, or may be a few such runs carried out within a relatively short calendar period. The objective generally emphasizes spatial distribution, and the characterization of the suite of parameters at a point in time. Examples are the Bureau of Economic Geology Submerged Lands Project and the U.S. Fish & Wildlife Service survey of sediment contaminants, in which each of a network of stations in the bay was sampled once to determine a basic suite of chemistry. A survey and a monitoring program share the feature that a suite of measurements is obtained that can be used to support different analyses and studies.

Finally, a *focused research* program is formulated to address a specific hypothesis, that in turn dictates the suite of measurements. As a practical matter, most focused research programs are limited in time, due to the nature of the funding process. While we differentiate between these three general strategies of sampling, it must be noted that there is considerable overlap: many monitoring programs provide data for research, many research programs comprise monitoring, and either can contain surveys as a part of the program. Some are ambiguous, such as the Corps of Engineers O&M sampling performed in association with dredging projects. These involve the occupation of a fixed network of stations, and with the same regularity as the need for maintenance dredging. But this is at such long intervals that one could argue that this program is in fact a series of surveys, rather than a monitoring program. Also, since the objective is to determine the effects of dredging on water and sediment quality, the program has many attributes of focused research. While there may not be clear differentiation between monitoring, surveys and focused research, the emphasis here is on the difference in philosophy underlying the program strategy, which can clarify the management and prosecution of various programs.

Extremely important to the present project are the presentation and dissemination of the basic data. Monitoring programs generally have provision for data storage and dissemination, nowadays digital. Surveys usually have some form of hard-copy presentation, and research programs may not publish or even preserve the basic measurements, but rather present analyzed or reduced data in a professional publication, or may extract the answer sought and file the data away.

Since this project seeks to compile and analyze a combined data set, machine processing is indispensable, and we therefore require all data in a machine-readable format. Where digital databases existed we sought copies from the managing agencies. In some instances, the digital record has been lost or destroyed. For a few of these we were able to recover the data record, in whole or in part. Where only hard copy or field notes existed, the data were keyboarded.

Even when a digital data base existed, transfer of the data for use in the present project was not necessarily straightforward. Just as the strategies of sampling programs may differ in basic objectives, so also may the strategies of digital data bases vary. There are two broad strategies observed in the maintenance of water and sediment quality data bases. First is the archiving and preservation of the basic data from a data-collection program. Second is support of various analytical procedures, such as basic statistical computations, or model validation. The same data base can achieve both strategies if either the measurements originate from a single program, or special care is taken in the development of the data base. If, however, data from more than one data-collection enterprise is involved, it is possible for the two objectives to become conflicting.

The distinction is exactly the same that was made in Chapter 4 between source files and derivative files. An *archival* data base is intended to preserve the measurements of a data-collection program, maximizing the information retained, without modifying or corrupting the data in any way. This includes compiling all ancillary data (such as time of sampling, observation of conditions, etc.), employing no units conversions (that cannot be simply reversed), and not pre-processing the basic measurements in any way (such as depth-compositing, time averaging, interpolating to standard space-time intervals, substituting values for BDL's, etc.). An *analytical* data base, in contrast, manipulates the basic measurements however necessary to facilitate the desired analyses. This may include averaging, smoothing, subsampling the basic data, or combining the measured data with measurements from programs of other agencies.

In the modern world of digital technology, all data-collection programs should have an archival data base. Examples include the U.S. Geological Survey STORET data base, which is the archival framework for streamflow and water quality data collected in the nation's streams and rivers by USGS, the EMAP data base maintained by EPA for archiving of data from the EMAP/REMAP project, and the data bases of Texas Department of Health and Texas Parks and Wildlife. Many agencies will also require various analyses based upon combined and processed data. The derivative data files developed for this CCBNEP Status & Trends project is an example of such an analytical data base. When an agency attempts to have a single data base serve both objectives, conflicts arise which limit the utility of the data, examples of which are cited below.

5.2.1 Principal Data-Collection Programs

Of central importance to Corpus Christi Bay are the existing monitoring programs, since these are the vehicles for continued, routine acquisition of data, and therefore form the backbone for determining the present water quality and any time trends. There are three major monitoring programs under way which contribute information on water and sediment quality of the bay, operated by the following agencies:

- Texas Natural Resource Conservation Commission
- Texas Parks & Wildlife Department
- Texas Department of Health

The Texas Natural Resource Conservation Commission Statewide Monitoring Network (SMN) is a principal continuing source of a broad spectrum of data. The SMN sampling program is a program of sampling at fixed stations at regular intervals, carried out by headquarters, field and/or District offices of the Texas Natural Resource Conservation Commission (TNRCC), and represents a continuity of activity through its predecessor agencies, the Texas Water Commission, Texas Department of Water Resources, and Texas Water Quality Board, back to approximately 1965. Generally, field parameters are obtained *in situ*, by means of electrometric probes or portable analytical kits, and water/sediment samples are shipped to external laboratories for analysis. (The laboratory used has varied over the years according to the parameter suite desired and to funds available for contracting. Past laboratories included the Texas State Department of Health, TNRCC/TWC Houston lab, Lower Colorado River Authority, Nueces County Health Department, and U.S. Corps of Engineers.) Parameters have been expanded from conventional variables in the early 1970's to trace constituents, pesticides and priority pollutants in recent years.

The term Statewide (a.k.a. Stream) Monitoring Network also refers to a data management system. The SMN data base is a digitized comprehensive data management program implemented on the TNRCC mainframe computer. The SMN data base includes all sampling activities of the Statewide Monitoring Network, as well as special studies (including microbiology and benthos) and Intensive Surveys. It also includes data from other agencies, notably Texas Water Development Board and the U.S. Geological Survey. There are over 1200 separate constituents with entries in the SMN data base, including water and sediment parameters, and biological parameters. In the five years since the Galveston Bay NEP Status & Trends project, the TNRCC has implemented sweeping changes in the structure and operation of this data base. Ward and Armstrong (1992b) were highly critical of the SMN protocols and data retrieval formats in place at that time, the latter being huge files of printer line images requiring magnetic tape for their transfer and manipulation. We are happy to report that these cumbersome procedures have been largely scrapped, in favor of a new system, presently under development, employing Internet file transfer protocols.

The Texas Parks & Wildlife Department or its predecessor agencies, the Texas Game and Fish Commission and the Texas Game, Fish and Oyster Commission, has monitored the fishery resources of the system for many years, and in association with this obtained a limited suite of water-quality variables. These tend to focus on estuarine habitat characteristics, e.g. salinity, dissolved oxygen, turbidity and temperature. While the range of variables is obviously much more limited than that of the TNRCC SMN, the temporal intensity of the program is much greater. The TPWD program obtains data somewhere in the system on virtually a daily basis, in contrast to the sampling interval of the SMN of one to several months. Further the spatial intensity is also greater. On the other hand, the TPWD samples a random network of stations, so there is no time continuity at a fixed point in the bay. The data is now entered into a digital data base at TPWD headquarters for detailed statistical analyses.

In order to regulate the harvesting of oysters in Corpus Christi Bay, the Seafood Safety Division (nee Shellfish Sanitation Division) of the Texas Department of Health (TDH) samples the bay at regular stations at varying temporal intensity, depending upon the season of year and upon the antecedent hydrological conditions. For the purpose of this program, the sampling is now limited to coliforms and a few associated hydrographic variables, salinity, temperature and pH. Like the TPWD, this program samples more intensely in space and time than the TNRCC SMN and has accumulated data from many years from Corpus Christi Bay. The collected data is maintained in a digital data base at TDH headquarters in Austin.

In addition, there are important recent or ongoing data collection programs in Corpus Christi Bay, also listed in Table 5-1, however these are not *monitoring* programs because they do not exhibit the regularity and time continuity implied by that term. One of the more important of these is the sampling performed by Galveston District Corps of Engineers in association with its Operations and Maintenance Program on navigation projects. This is intense sampling emphasizing sediment quality that is performed in association with dredging activities. The sampling interval is therefore dictated by the condition of the channel, i.e. sediment accumulation, and may be as long as several years. The Corps data program has been subdivided in Table 5-1 according to the suite of parameters obtained. Generally, there has been an evolution from an emphasis on conventional chemistry and metals to specific hydrocarbons.

Of the historical programs available, there are several which are noteworthy. The Texas Game, Fish and Oyster Commission (TGFOC) is the predecessor organization of the present TPWD, and has sampled the system on a routine basis back to the early 1950's, and on an occasional basis back to the Nineteenth Century. Data antedating 1975 is extremely inaccessible. Most of it is stored as hard-copy records (i.e., original field sheets) in a state warehouse in Olmeto. Unfortunately, the resources of this project did not permit the major effort that would be required to exhume and keyboard this old data. We did have access to some of this data copied in special project reports of the TGFOC, and extracted from older annual reports of the Coastal Fisheries Branch. This data was digitized and incorporated into the data base. It is urgent that an effort be made to recover the original data holdings and render them in a digital form for future researchers. (See Section 5.4.2, below.)

One of the major historical studies developed from the operation of the Ocean Science and Engineering Laboratory of Southwest Research Institute (SWRI) in Corpus Christi from the late-1960's to the mid-1970's. In concert with this lab, a routine monitoring program in Corpus Christi Bay was operated from 1970-75 with hiatuses due to shortage of funds. Also, several special-purpose studies (an example of focused research projects) were performed by the labs under sponsorship of regional agencies and industries. When SWRI closed the lab, all of its data holdings were removed and probably destroyed. SWRI would not disclose to these investigators whether or where the data presently exists. Fortunately, John Buckner of the Coastal Bend Council of Governments made available to this project his considerable archive of data from the system, that included reports from SWRI reproducing most of the measurements. These data were keyboarded for this project.

Another noteworthy program is the Submerged Lands Study of the University of Texas Bureau of Economic Geology, sponsored by the Texas General Land Office. This program, which focused entirely upon sediment, falls into the category of a survey, because it involved one-time only sampling. However, it is the only data set extant which sampled the *entirety* of Corpus Christi Bay at a uniform station distribution (1-mile), irrespective of the location of shoals, channels, navigation aids and reefs (which tend to spatially bias most measurements from the system).

5.2.2 Summary of the data base

The data programs of Table 5-1 formed the basis for the analysis reported here. Most of these programs, it will be noted, are small-scale focused research activities, though most of the data is dominated by the few large-scale programs summarized above. The approach of this project was to combine and merge these programs to synthesize a more comprehensive data base for the system. Details on the data sets of these individual programs are given in the companion data base report (Ward and Armstrong, 1997a), along with any problems encountered in the data and how those problems were resolved (or reconciled). Particular note should be made of the programs which were keyboarded into a digital format for this project. As noted earlier, this digital data set, which is capable of much more analysis than it is subjected to here, is considered one of the chief products of this project.

The data bases for water quality and for sediment quality are summarized on a baywide basis in Tables 5-2 and 5-3, resp. The only values screened from analysis at this stage are those that exceed the (rather liberal) bounds on variables given in Table 4-1. (Parameter names are abbreviated for compactness; their definitions are given in Table 2-1.) These tables are information-dense, and largely self-explanatory. These tables do provide a ready index to the relative intensity with which different variables have been measured, and the extant period of record. Because of the large spatio-temporal variability in most of these parameters, the baywide means have little significance; however they are useful in typifying the magnitudes of the different variables (provided the spurious values do not seriously corrupt the mean).

Because only the bounds of Table 4-1 are used to screen data, the summary of data in Tables 5-2 and 5-3 reflects some probably-incorrect values. There is such a large range of possible values in the Corpus Christi system that the demarcation between probable and improbable is not at all certain: the 10 ppb values for water-phase 2,4,5-T and Silvex, or the 640 μg/kg of sediment BaP may be more unlikely than unusual, but we are hesitant to dismiss them on strictly an *a priori* basis.

Table 5-2
Summary statistics for water analytes for entire CCBNEP study area

*Parameter**	*No.of obs*	*Avg >DL*	*St dev >DL*	*No. > DLs*	*% > DLs*	*Min*	*date*	*Min >0*	*date*	*Max*	*date*	*Average BDL=0*	*with =DL*
Conventional parameters													
WQALK	2506	154	66	2506	100	0	770202	11	930511	830	910808	154	154
WQAMMN	9221	0.177	0.88	8138	88	0	711111	0.00014	930415	30	750422	0.156	0.163
WQBOD5	2069	2.37	2.8	2049	99	0.19	720707	0.19	720707	90	701119	2.34	2.35
WQCHLA	5492	13.2	31	4823	88	0	730615	0.1	740417	1100	880412	11.6	12.1
WQCHLB	62	2.1	1.5	62	100	1	720919	1	720919	10	751030	2.1	2.1
WQDO	49280	7.47	2.2	49280	100	0	681121	0.05	710825	25	890417	7.47	7.47
WQFCOLI	6010	569	13000	5219	87	0	740612	1	720918	340000	761104	495	499
WQKJLN	2042	1.02	1.2	1981	97	0	750627	0.01	760212	34	841018	0.99	0.995
WQNO2N	7239	0.0196	0.1	6365	88	0	711105	0.00014	880707	4.8	680326	0.0173	0.0205
WQNO3N	9136	0.148	0.79	8021	88	0	710609	0.00014	880714	19	720817	0.13	0.134
WQO&G	115	6.94	8.9	71	62	0.2	870331	0.2	870331	38	751209	4.28	4.67
WQOPD	4425	0.204	0.77	4057	92	0	720918	0.003	860102	12	860428	0.187	0.188
WQOPO4	4316	0.428	1.7	4003	93	9.5E-4	871020	0.00095	871020	27	741218	0.397	0.399
WQORGN	1402	0.556	0.34	1402	100	0	750827	0.005	740416	5	760128	0.556	0.556
WQPH	22586	8.15	0.36	22586	100	4	850930	4	850930	10	730518	8.15	8.15
WQPHEO	3841	5.39	9.8	2480	65	0	770719	0.1	871210	280	930427	3.48	5.05
WQSAL	58109	26.1	11	58108	100	0	730615	0.0036	920728	300	850922	26.1	26.1
WQSECCHI	4732	0.772	0.5	4664	99	0	800701	0.0004	830512	7.7	750604	0.761	0.765
WQSIO2	3726	4.06	3.6	3726	100	0	720918	0.015	880510	32	690325	4.06	4.06
WQSO4	4581	1740	1100	4569	100	1	740531	1	740531	4900	900807	1730	1730
WQTCOLI	3250	442	7200	2906	89	0	740612	0.01	821118	240000	610927	395	396
WQTEMP	58085	23.1	6	58084	100	0.1	731029	0.1	731029	39	850918	23.1	23.1
WQTOC	4741	10.6	9.3	4157	88	0	860507	0.24	720328	84	740417	9.34	9.79
WQTOTP	6627	0.241	0.85	6593	99	0	710608	0.006	750106	28	860428	0.24	0.24

(continued)

Table 5-2
(continued)

Parameter	No.of obs	Avg >DL	St dev >DL	No. > DLs	% > DLs	Min	date	Min >0	date	Max	date	Average BDL=0	with =DL
Conventional parameters continued													
WQTPO4	2259	0.972	2.8	2221	98	0	760729	0.02	750106	34	820505	0.956	0.957
WQTRANS	3536	62	22	3536	100	0	701212	1	700610	100	720514	62	62
WQTSS	2964	60.8	110	2904	98	0	860507	0.27	790810	2000	710903	59.5	59.7
WQTURBJ	12709	48.1	61	12627	99	0	730815	1	720724	1000	830420	47.8	48.1
WQTURBN	11851	16.6	26	11851	100	0	870518	0.8	870110	1000	870401	16.6	16.6
WQVOLS	16	4380	250	16	100	3900	751209	3900	751209	4900	751209	4380	4380
WQVSS	3233	16.8	21	2718	84	0	720115	0.1	910116	320	900213	14.1	15.6
Metals													
WQMETAGD	95	1	1.5	23	24	0	791017	1	880809	7	860903	0.242	1.81
WQMETAGT	191	51.6	120	40	21	0.36	930109	0.36	930109	790	770105	10.8	16.6
WQMETASD	151	6.04	5.8	143	95	0	720918	1	891017	31	820525	5.72	6.14
WQMETAST	460	10.8	19	168	37	0.35	771228	0.35	771228	220	770105	3.96	6.29
WQMETBAD	134	253	170	89	66	18	761025	18	761025	620	830824	168	221
WQMETBAT	59	50.6	17	58	98	11	940804	11	940804	92	940316	49.8	49.8
WQMETBD	93	1690	1200	93	100	0.21	680226	0.21	680226	5000	711109	1690	1690
WQMETBT	20	2210	2000	20	100	4.4	770727	4.4	770727	5000	740723	2210	2210
WQMETCDD	158	1.45	4.6	103	65	0	720918	0.94	910808	30	801022	0.943	1.32
WQMETCDT	473	16.8	20	74	16	0	741017	1.3	920422	70	930216	2.63	4.95
WQMETCOD	52	0.0385	0.19	52	100	0	731015	1	731016	1	731016	0.0385	0.0385
WQMETCOT	12	1.58	2	12	100	0	741017	3	741024	5	741024	1.58	1.58
WQMETCRD	144	8.07	15	94	65	0	720918	1	840409	80	850122	5.27	7.39

(continued)

Table 5-2
(continued)

Parameter	No.of obs	Avg >DL	St dev >DL	No. > DLs	% > DLs	Min	date	Min >0	date	Max	date	Average BDL=0	with =DL
Metals continued													
WQMETCRT	463	44.9	32	76	16	1.2	940316	1.2	940316	100	740731	7.36	16.7
WQMETCUD	167	7.63	11	150	90	0	791017	1	820210	80	801022	6.85	7.19
WQMETCUT	472	26.1	37	162	34	0.6	720321	0.6	720321	240	761101	8.96	11.8
WQMETFED	83	72.8	55	83	100	0	720918	10	741017	210	810324	72.8	72.8
WQMETFET	191	523	700	186	97	0	750917	1	750814	5000	830810	509	510
WQMETHGD	70	0.348	0.56	29	41	0	811020	0.1	811119	3	871013	0.144	0.229
WQMETHGT	457	0.832	1.4	69	15	0.1	811119	0.1	811119	11	770216	0.126	0.427
WQMETMND	87	40	31	87	100	0	720918	10	811209	120	741023	40	40
WQMETMNT	136	99.2	340	127	93	6	860528	6	860528	3700	820819	92.7	93.5
WQMETNID	108	2.6	5.6	80	74	0	751103	1	840117	31	910808	1.93	3.16
WQMETNIT	123	28.4	48	50	41	1	811119	1	811119	280	740712	11.6	20.3
WQMETPBD	154	3.24	5.7	103	67	0	720918	1	840117	36	751103	2.17	3.36
WQMETPBT	465	45.6	46	111	24	0	741017	1	741017	200	830810	10.9	22.2
WQMETSED	91	0.831	0.46	29	32	0	791016	1	811119	2	840709	0.265	1.26
WQMETSET	297	6.68	10	37	12	1	811119	1	811119	49	930109	0.832	2.64
WQMETSRD	21	147	250	21	100	0.09	670927	0.09	670927	860	731017	147	147
WQMETZND	156	45.6	89	141	90	0	720918	2	720928	820	731017	41.2	43.1
WQMETZNT	469	88.3	170	263	56	1	720321	1	720321	2000	771228	49.5	53.7
Organics													
WQ-245T	116	0.122	1.1	83	72	0	750623	0.01	751103	10	760922	0.0875	1.8
WQ-24D	115	7.6E-3	0.029	82	71	0	750623	0.01	741024	0.2	711105	5.4E-3	7.67
WQ-ABHC	62	0	0	0	0							0	2.49

(continued)

Table 5-2
(continued)

Parameter	No.of obs	Avg >DL	St dev >DL	No. > DLs	% > DLs	Min	date	Min >0	date	Max	date	Average BDL=0	with =DL
Organics continued													
WQ-ACEN	247	0	0	0	0							0	2.62
WQ-ACENA	12	0	0	0	0							0	2.5
WQ-ALDR	107	0.0002	0.0014	50	47	0	720918	0.01	830629	0.01	830629	9.4E-5	0.864
WQ-ANTHR	19	4.15	2.6	2	11	1.6	840508	1.6	840508	6.7	840508	0.437	1.12
WQ-BNZA	243	0	0	0	0							0	1.29
WQ-BNZAA	12	0	0	0	0							0	1
WQ-BNZB	12	0	0	0	0							0	0.1
WQ-BNZGP	12	0	0	0	0							0	0.1
WQ-BNZK	12	0	0	0	0							0	0.1
WQ-CHLR	325	0.0049	0.02	49	15	0	720918	0.05	891219	0.1	830629	7.4E-4	1.45
WQ-CHLRC	5	0	0	0	0							0	0.185
WQ-CHRYS	16	0	0	0	0							0	0.625
WQ-DBANE	12	0	0	0	0							0	0.5
WQ-DDD	50	0.0002	0.0014	50	100	0	720918	0.01	830629	0.01	830629	0.0002	0.0002
WQ-DDE	50	0.0002	0.0014	50	100	0	720918	0.01	830629	0.01	830629	0.0002	0.0002
WQ-DDT	50	0.0004	0.002	50	100	0	720918	0.01	830629	0.01	830629	0.0004	0.0004
WQ-DIAZ	46	0.0744	0.41	32	70	0	720918	0.01	731015	2.4	871104	0.0517	0.15
WQ-DIEL	107	0.0002	0.0014	50	47	0	720918	0.01	830629	0.01	830629	9.4E-5	0.86
WQ-ENDO	20	1.7E-3	0.0037	6	30	0	791017	0.01	830629	0.01	830629	0.0005	4.51
WQ-ENDR	75	0.0002	0.0014	50	67	0	720918	0.01	830629	0.01	830629	1.3E-3	1.24
WQ-FLRA	219	0	0	0	0							0	0.5
WQ-FLRN	20	2.1	1.1	3	15	0.7	840508	0.7	840508	3.5	840508	0.315	0.865
WQ-HEPT	128	8.9E-4	0.0054	56	44	0	750623	0.01	830629	0.04	760922	4E-4	0.715

(continued)

Table 5-2
(continued)

Parameter	No.of obs	Avg >DL	St dev >DL	No. > DLs	% > DLs	Min	date	Min >0	date	Max	date	Average BDL=0	with =DL
Organics continued													
WQ-HEPX	55	0.0035	0.013	20	36	0	750623	0.01	830629	0.06	760922	0.00127	0.0525
WQ-HEXA	10	0	0	0	0							0	0.0198
WQ-I123P	12	0	0	0	0							0	0.5
WQ-LIND	86	3.3E-3	0.0094	9	10	0	720918	0.03	760922	0.03	760922	0.000349	1.07
WQ-MALA	87	0.0271	0.19	52	60	0	750623	0.01	830629	1.4	760922	0.0162	0.315
WQ-MTHP	40	0	0	31	77	0	720918					0	0.0517
WQ-MTHX	57	0.0925	0.3	12	21	0	750623	0.01	830629	1.1	760922	0.0195	2.01
WQ-NAPT	247	7.4	0	1	0	7.4	840508	7.4	840508	7.4	840508	0.03	2.64
WQ-ODDT	5	0	0	0	0							0	0.126
WQ-PAH	219	0	0	0	0							0	3.77
WQ-PARA	72	0.0135	0.081	37	51	0	750623	0.5	760922	0.5	760922	0.00694	0.173
WQ-PCB	342	0.126	0.38	58	17	0	750623	0.1	720918	2	740505	0.0213	0.502
WQ-PCP	31	0	0	0	0							0	33.8
WQ-PDDD	41	0	0	0	0							0	2.28
WQ-PDDE	41	0	0	0	0							0	2.21
WQ-PDDT	265	7.78	0.88	8	3	6.2	930216	6.2	930216	8.8	930216	0.235	0.623
WQ-PHNAN	16	0	0	0	0							0	1
WQ-PYRN	12	0	0	0	0							0	0.5
WQ-SLVX	112	0.121	1.1	83	74	0	750623	0.01	830629	10	760922	0.0895	1.67
WQ-TOXA	299	0.0435	0.2	23	8	0	741017	1	830629	1	830629	0.00334	0.954

* Abbreviations & units defined in Table 2-1

Table 5-3
Summary statistics for sediment analytes for entire CCBNEP study area

*Parameter**	*No. of obs*	*Avg >DL*	*Stdev >DL*	*No.> DLs*	*%> DLs*	*Min*	*date*	*Min >0*	*date*	*Max*	*date*	*Average BDL=0*	*with =DL*
Conventional parameters													
SEDAMMN	61	89.1	66	38	62	3.1	840508	3.1	840508	280	770621	55.5	55.9
SEDCYAN	38	0.629	0.34	38	100	0.2	760421	0.2	760421	1.8	760420	0.629	0.629
SEDKJLN	264	988	900	259	98	20	731029	20	731029	6800	880426	969	969
SEDO&G	438	893	1100	401	92	0	720716	5	880127	9600	840508	817	823
SEDORGN	11	834	170	11	100	500	770621	500	770621	1200	770621	834	834
SEDTOC	1251	7.26	6.7	1180	94	0.0031	950111	0.0031	950111	64	760924	6.85	6.85
SEDTOTP	267	573	950	266	100	0.1	740723	0.1	740723	8900	771229	571	572
SEDVOLS	270	55600	39000	270	100	0	740723	1	740723	190000	830505	55600	55600
Metals													
SEDMETAG	657	1.57	7.4	165	25	0.033	930109	0.033	930109	90	860527	0.395	2.44
SEDMETAL	273	11600	8800	273	100	320	880615	320	880615	47000	880615	11600	11600
SEDMETAS	950	4.31	5.3	740	78	0	720918	0.06	780607	75	770105	3.36	3.49
SEDMETB	578	37.3	31	554	96	0.69	800825	0.69	800825	340	880615	35.8	35.9
SEDMETBA	845	345	270	842	100	3	810831	3	810831	2700	760331	344	344
SEDMETCD	960	5.61	12	453	47	0	731015	0.02	920729	110	760406	2.65	2.81
SEDMETCO	28	6.98	3.6	28	100	1.2	720918	1.2	720918	11	731017	6.98	6.98
SEDMETCR	1243	19.2	22	1143	92	0	840623	0.3	800819	260	720716	17.6	18
SEDMETCU	1253	11.1	16	1147	92	0.3	880615	0.3	880615	220	940316	10.2	10.4
SEDMETFE	674	11900	8500	674	100	490	880615	490	880615	39000	890902	11900	11900
SEDMETHG	944	0.294	0.81	573	61	0	750129	0.01	741220	8.5	750115	0.179	0.246
SEDMETMN	956	276	180	956	100	8.6	731029	8.6	731029	1200	860814	276	276

(continued)

Table 5-3
(continued)

*Parameter**	*No. of obs*	*Avg >DL*	*Stdev >DL*	*No.> DLs*	*%> DLs*	*Min*	*date*	*Min >0*	*date*	*Max*	*date*	*Average BDL=0*	*with =DL*
						Metals continued							
SEDMETNI	1223	8.8	6.7	1028	84	0.2	760902	0.2	760902	80	770105	7.39	8.35
SEDMETPB	1095	27.1	48	868	79	1	860507	1	860507	450	720716	21.5	22.5
SEDMETSE	656	0.675	0.86	173	26	0.067	930109	0.067	930109	7.2	890424	0.178	0.92
SEDMETSR	584	299	410	584	100	13	880615	13	880615	3500	760921	299	299
SEDMETZN	1286	180	630	1175	91	0.6	760902	0.6	760902	6200	720716	165	166
						Organics							
SED-245T	71	0.25	0.35	6	8	0	741024	0.7	720918	0.8	720918	0.0211	26.3
SED-24D	70	1.06	1.3	5	7	0	741023	2.6	720918	2.7	720918	0.0757	63.6
SED-ABHC	54	0	0	0	0							0	40
SED-ACEN	309	2.25	2.8	23	7	0.11	910805	0.11	910805	11	910805	0.168	33.5
SED-ACNY	58	1.9	3.4	22	38	0	860715	0.08	910805	15	880715	0.721	10.4
SED-ALDR	340	0.115	0.31	95	28	0	731015	0.01	860715	2	880715	0.0321	6.75
SED-ANTH	146	25.7	60	36	25	0	880715	0.2	930720	350	860715	6.35	21.1
SD-BNZAA	60	27.9	34	33	55	10	880615	10	880615	120	880615	15.4	25.7
SED-BNZA	369	52.7	100	74	20	0.06	920729	0.06	920729	640	940316	10.6	30.2
SD-BNZB	12	85.3	92	3	25	16	910822	16	910822	220	910822	21.3	41.5
SD-BNZGP	12	24	16	2	17	8	910822	8	910822	40	910822	4	24.3
SD-BNZK	12	48	51	3	25	8	910822	8	910822	120	910822	12	32.2
SED-CHLR	513	1.28	4.2	75	15	0	731015	0.01	930720	20	760831	0.188	17.8
SD-CHLRC	50	6	1	2	4	5	880523	5	880523	7	880525	0.24	3.69
SED-CHRY	143	33.1	65	70	49	0	880615	0.5	930720	490	880615	16.2	28.1

(continued)

Table 5-3
(continued)

*Parameter**	*No. of obs*	*Avg >DL*	*Stdev >DL*	*No.> DLs*	*%> DLs*	*Min*	*date*	*Min >0*	*date*	*Max*	*date*	*Average BDL=0*	*with =DL*
						Organics continued							
SD-DBANE	12	16	0	1	8	16	910822	16	910822	16	910822	1.33	23.8
SED-DDD	195	0.455	1.2	77	39	0	751030	0.1	791016	7.6	711105	0.179	3.08
SED-DDE	198	2.01	6.7	83	42	0	751030	0.02	741017	55	830712	0.845	2.61
SED-DDT	215	0.666	3	96	45	0	731015	0.016	920729	28	750414	0.298	3.21
SED-DIAZ	157	0.375	1.3	40	25	0	751103	5	760831	5	760831	0.0955	4.69
SED-DIEL	346	0.256	0.7	80	23	0	731015	0.0023	920729	3	760831	0.0593	7.2
SED-ENDO	22	0.15	0.05	2	9	0.1	830629	0.1	830629	0.2	940724	0.0136	54.9
SED-ENDR	252	0.282	0.83	61	24	0	731015	0.1	830629	3.6	830331	0.0684	10.2
SED-FLRA	347	48.5	84	95	27	0	840930	1	930720	450	880615	13.3	32.6
SD-FLRN	147	7.38	8.6	29	20	0.2	930720	0.2	930720	30	880615	1.46	13.7
SED-HEPT	274	0.0846	0.23	54	20	0	751030	0.00059	920729	1	760831	0.0167	4.86
SED-HEPX	314	0.0922	0.23	86	27	0	731015	0.01	880715	1.1	870715	0.0253	3.54
SED-HEXA	57	32	0	1	2	32	860716	32	860716	32	860716	0.561	15.3
SD-I123P	12	40	0	1	8	40	910822	40	910822	40	910822	3.33	25.8
SED-LIND	310	0.119	0.28	39	13	0	750604	0.01	930722	1	760831	0.015	7.51
SED-MALA	62	0	0	20	32	0	751103					0	5.32
SED-MTHP	156	0.596	2.2	47	30	0	751103	5	760831	13	770914	0.179	2.67
SED-MTHX	158	1.95	5.9	31	20	0	750604	0.1	830629	20	760831	0.384	18.3
SED-NAPT	363	11.3	26	29	8	0.54	920729	0.54	920729	140	860715	0.905	39
SED-ODDD	49	0	0	0	0							0	3.56
SED-ODDE	50	0	0	0	0							0	2.23
SED-ODDT	50	0	0	0	0							0	3.55

(continued)

Table 5-3
(continued)

Parameter*	No. of obs	Avg >DL	Stdev >DL	No.> DLs	%> DLs	Min	date	Min >0	date	Max	date	Average BDL=0	with =DL
						Organics continued							
SED-PAH	237	226	670	27	11	0.2	870331	0.2	870331	3500	940316	25.8	26.2
SED-PARA	148	0.484	1.5	31	21	0	750604	5	760831	5	760831	0.101	3.76
SED-PCB	520	40.3	79	117	22	0	751103	0.1	910806	420	790808	9.06	25.4
SED-PCP	68	0	0	0	0							0	347
SED-PDDD	216	0.352	1.1	47	22	0	870715	0.0035	920729	7	900715	0.0766	11.2
SED-PDDE	216	0.818	2.1	53	25	0	870715	0.0067	920729	12	900715	0.201	10.7
SED-PDDT	437	0.704	1.5	45	10	0	870715	0.01	860715	7	900715	0.0725	7.46
SED-PERY	65	20.7	18	46	71	0	890902	0.21	920729	63	840930	14.6	20.8
SD-PHNAN	140	61.7	170	39	28	0.6	930720	0.6	930720	1000	860715	17.2	32.4
SD-PYRN	125	55.7	82	69	55	0.7	900715	0.7	900715	360	880615	30.8	44.6
SED-SLVX	111	1.81	5.7	34	31	0	750604	0.7	720918	20	760831	0.554	10.9
SED-TBT	18	5.93	9.1	12	67	0.07	930720	0.07	930720	29	930722	3.95	5.22
SED-TOXA	497	3.46	12	46	9	0	751103	10	830629	50	760631	0.322	41.1

* Abbreviations & units defined in Table 2-1

The highest value of chlorophyll-a of 1100, measured by the TNRCC, is so large that one might suspect a data-point error. One of the highest values of DO was an unrealistic 22.5. Both of these measurements might have been legitimately screened out. But when one reflects that they were both obtained from the *same* water sample, the possibility is admitted that, though unlikely, they may be real.

There are three different averages presented in these tables. The first is the average of those measurements excluding the censored data (i.e., below detection limits, BDL's). In the last two columns, two additional averages are presented, the first based upon assuming all BDL data to be zero (0), the second assuming all such censored data to be equal to the specified detection limit. (See Section 4.3.2 for discussion.) Clearly, for those parameters which have not been censored, the three averages are equal.

One problem facing a user of this data set is the interpretation of zero values. We are concerned that some of the zero values may in fact have originated from blank entries (i.e., no measurement) that in the process of agency transcription and digitization were replaced with a zero. This was especially a problem in the early days (the 1960's) of digital processing, in which most of the software was special-purpose FORTRAN codes. The FORTRAN editing specifications will interpret a blank field as a zero, unless special (and complicated) provision is made in the code, provisions which were usually not indulged in by programmers of the 1960's. While this anomaly was probably provided for in the first stage of data processing, as the data was passed through additional FORTRAN codes, we suspect that the differentiation between a blank field and a zero was lost. Parameters such as TSS and water-phase nitrogens exhibit these too-frequent values of zero.

Another problem with zeroes in the data record arises from measurements that were below detection limits. For many of the trace parameters, such as metals and pesticides, the lower range of measurement in the data base is delimited by the detection limits of the procedure. Despite this, several of the data sets, especially those from the Texas Water Development Board, include zero values especially for data from the early 1970's. We suspect that these were introduced for analytical purposes to replace the censoring statement "below detection limits" (BDL). Some may have also been blanks misinterpreted as zeroes, as described above. The data analyzed to produce Table 5-2 and 5-3 have not been screened for such anomalies. For trace constituents whose records include zero values, the zeroes are interpreted in these averages as being *above* the detection limit. In the detailed analyses of Chapter 6 and 7, these anomalous zero values were replaced with an estimated detection limit, see Section 4.2.4.

It should be noted that some of the EPA priority pollutants do not appear in Tables 5-2 and 5-3. Very few measurements have been made in Corpus Christi Bay of most of the priority pollutants. In some instances, there may be a scattering of measurements, such as from the recent OxyChem program (Project Code 15 in Table 5-1), but not enough to use in any meaningful way in a status-and-trends analysis. For example, there were two measurements of water-phase Endrin aldehyde from the entire Corpus Christi Bay system. Similarly, most of the individual PAH's were represented only by a handful of data, mainly from the EPA EMAP program or from the NOS Status & Trends Project. Those variables for which the sample base is totally lacking or inadequate are excluded from these tables and from this analysis. For those parameters for which there is at least a minimum analyzable data base, most of those measurements are below detection limits (BDL), as indicated in these tables. (There is no *a priori* means of specifying how many measurements comprise a "minimum analyzable data base." We have arbitrarily required at least 10, taken over a period exceeding one year. Given the uncertainty in the analytical measurement itself, to say nothing of variability in the bay, the statistical resolving power of such a small data set is feeble.)

Figures 5-1 *et seq.* display graphically the sampling intensity throughout the bay for the more important of the water and sediment parameters. Sampling intensity is measured by the number of observations within each Hydrographic Segment (see Section 3.3.3) for the period of record. The amount of data available is strongly dependent upon the parameter. Salinity and temperature, referred to as "hydrographic parameters," are easily measured in the field, and so the data holdings are large, approaching 60,000 independent measurements. Dissolved oxygen and pH are practically in the same category, as electrometric probes suitable for field use became common in the late 1960's. Conventional water-quality chemical parameters, for which a water sample is returned to a laboratory and subjected to straightforward wet-chemistry techniques (including filtration, ignition, titration, incubation, etc.), are represented by BOD (Figs. 5-19 through 5-24), ammonia, total phosphorus, total suspended solids, and fecal coliforms (Figs. 5-43 through 5-48). As would be expected, the data holdings for these parameters are about an order-of-magnitude less than the hydrographic parameters, and significant areas of the system have never been sampled. Cadmium and the proxy data file for DDT are presented as representative of the metals and organics parameters, as the data holdings are largest for these parameters. It is apparent that, in comparison to the conventional parameters, only a handful of measurements of metals and organics are available.

Several observations may be made about these figures. First, what might appear to be a large number of historical samples for a given parameter on a baywide basis, from Tables 5-2 and 5-3, is shown to be quite modest—even inadequate—when related to specific areas of the bay. Second, it is apparent at once from Figs. 5-1 *et seq.* that sampling intensity is highly heterogeneous in space, some areas of the bay having been subjected to relatively frequent sampling, and some rarely sampled. There is a particular bias, as might be expected, for the main navigation channels (ease of access and of positioning) and for those areas with historical pollution problems such as the Inner Harbor. The period of record generally ranges over many years so the number of samples *per year* is a considerably smaller number.

There is roughly an order of magnitude less sediment data from Corpus Christi Bay than water quality data. This does not, however, imply that the data base for sediment is an order of magnitude less useful. On the contrary, sediment is much less mobile than water and sedimentary processes proceed much more slowly than those in the aqueous phase. For parameters bound to sediment, which include most of the nutrients and practically all of the trace organics and metals, sediment can be considered a long-term integrator of water-phase concentrations. Also, concentrations accumulate in sediment, and because they are larger than those in the overlying water, can be measured with a higher degree of confidence. For these reasons, generally less data per unit time is considered to be necessary to characterize sediment concentrations compared to those in water. On the other hand, sediment is usually more heterogeneous in space than water, so the spatial sampling density can be a limitation to analysis of sediment concentrations.

Density of sediment samples is exemplified by sediment oil & grease in Figs. 5-61 through 5-66, zinc in Figs. 5-67 through 5-72, and the proxied DDT in Figs. 5-73 through 78, as indicators of conventional sediment chemistry, sediment metals and sediment organics, respectively. Metals have generally received the greatest attention by the monitoring projects, so almost all areas of the system have been sampled at least a few times for some of the metals. For conventional parameters and organics, many areas of the bay have never been sampled. More so than for water samples, there is an investment of time and equipment in obtaining a sediment sample, so large suites of parameters are typically run for each sample. This means that there is a high correlation in the sampling density among most metals parameters, and among most volatile organics. Therefore, these three sets of figures are very representative of their respective classes of analytes.

5.3 Data Base Deficiencies

The general adequacy of coverage in Corpus Christi Bay for each parameter is an important dimension of this compilation. This includes identification of data gaps in the record and their associated implications for application of the record to long-term trend analysis (seasonal bias is a common problem, for example). It also includes procedural shortcomings that prejudice or limit the applicability of data collected. These are discussed in Section 5.3.1 for specific parameters and classes of parameters.

Even after the efforts of this project to locate, compile and synthesize a "complete" digital data base, the period of record extends back only to about 1965 for conventional indicators, and 1975 for metals and organics. For some traditional parameters, and for certain areas of the bay, e.g., the Inner Harbor, the record can be extended back to the late 1950's. Earlier than this, what data still survive are sporadic in time, e.g. some temperature and salinity measurements in the early 1950's, a few from the 1940's, one set of dissolved oxygen measurements from Laguna Madre in 1948, etc. This earlier data is generally inadequate in space and time continuity for any reliable characterization or trends analysis: data prior to 1950 are therefore excluded from analyses of the data base.

At least for standard hydrographic parameters and some of the traditional quality indicators, much more data than this has been taken from the Corpus Christi Bay system. The data base available to this project represents a fraction of the data

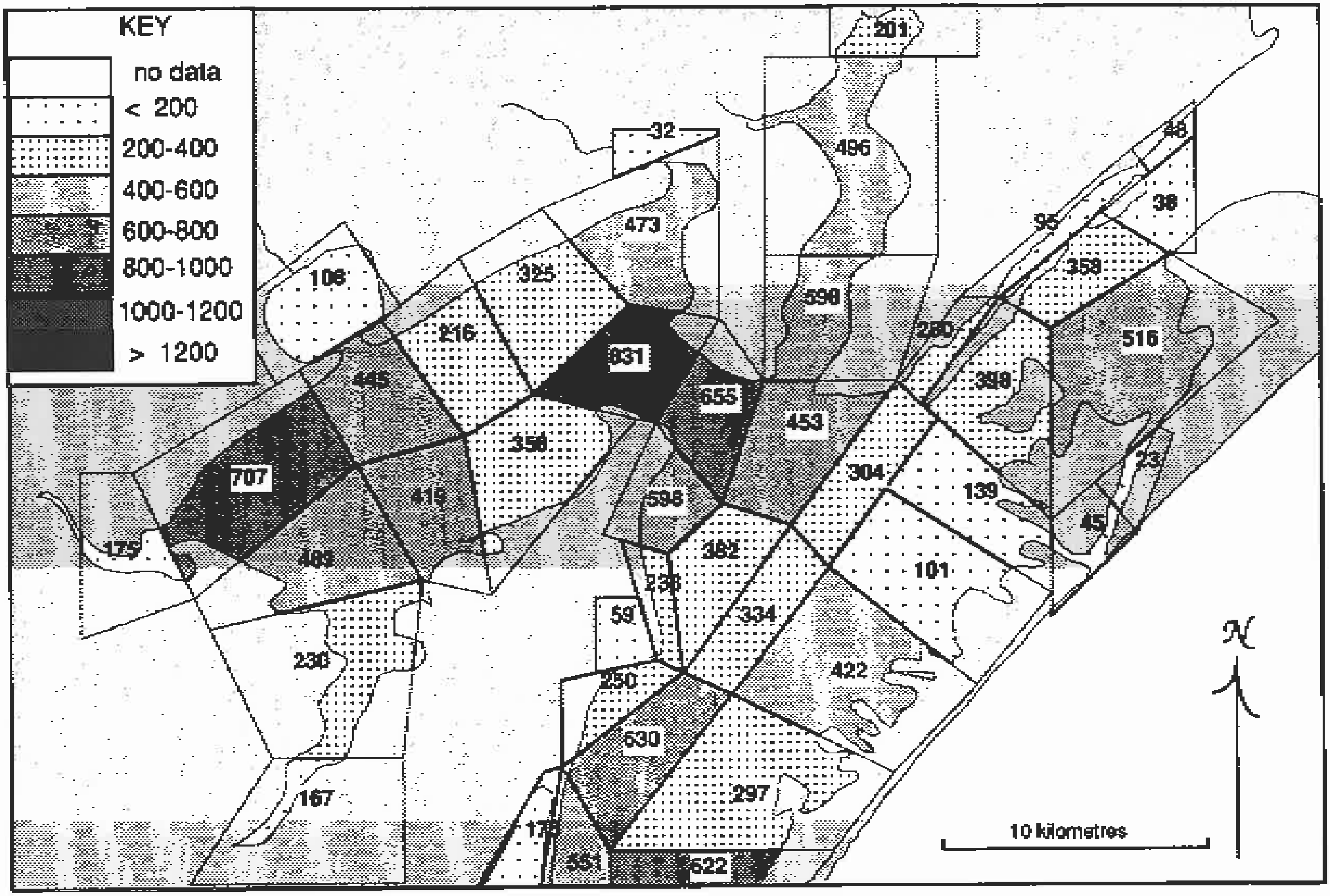


Figure 5-1. Sampling density (total number of measurements) of WQSAL for Aransas-Copano system

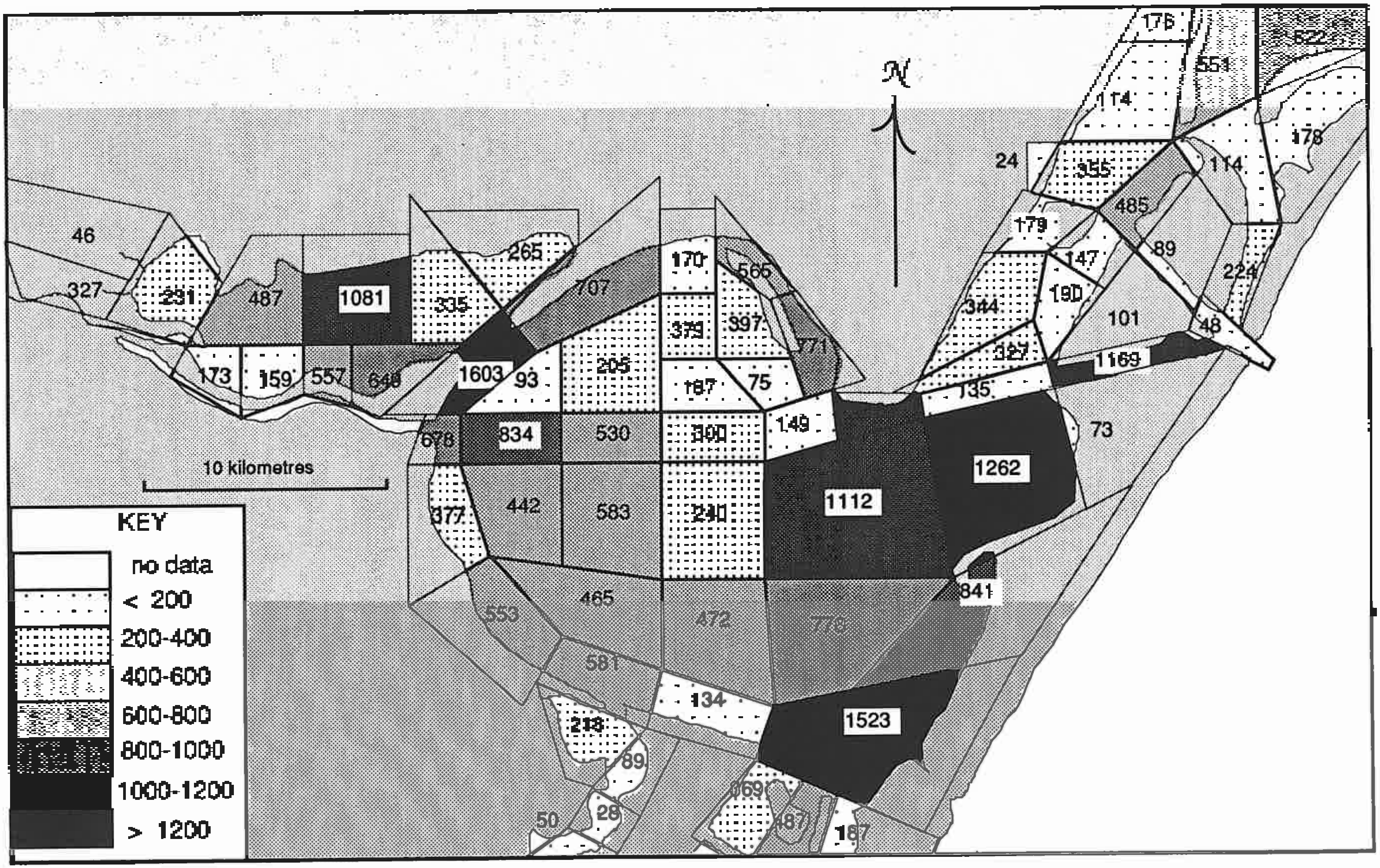


Figure 5-2. Sampling density (total number of measurements) of WQSAL for Corpus Christi system

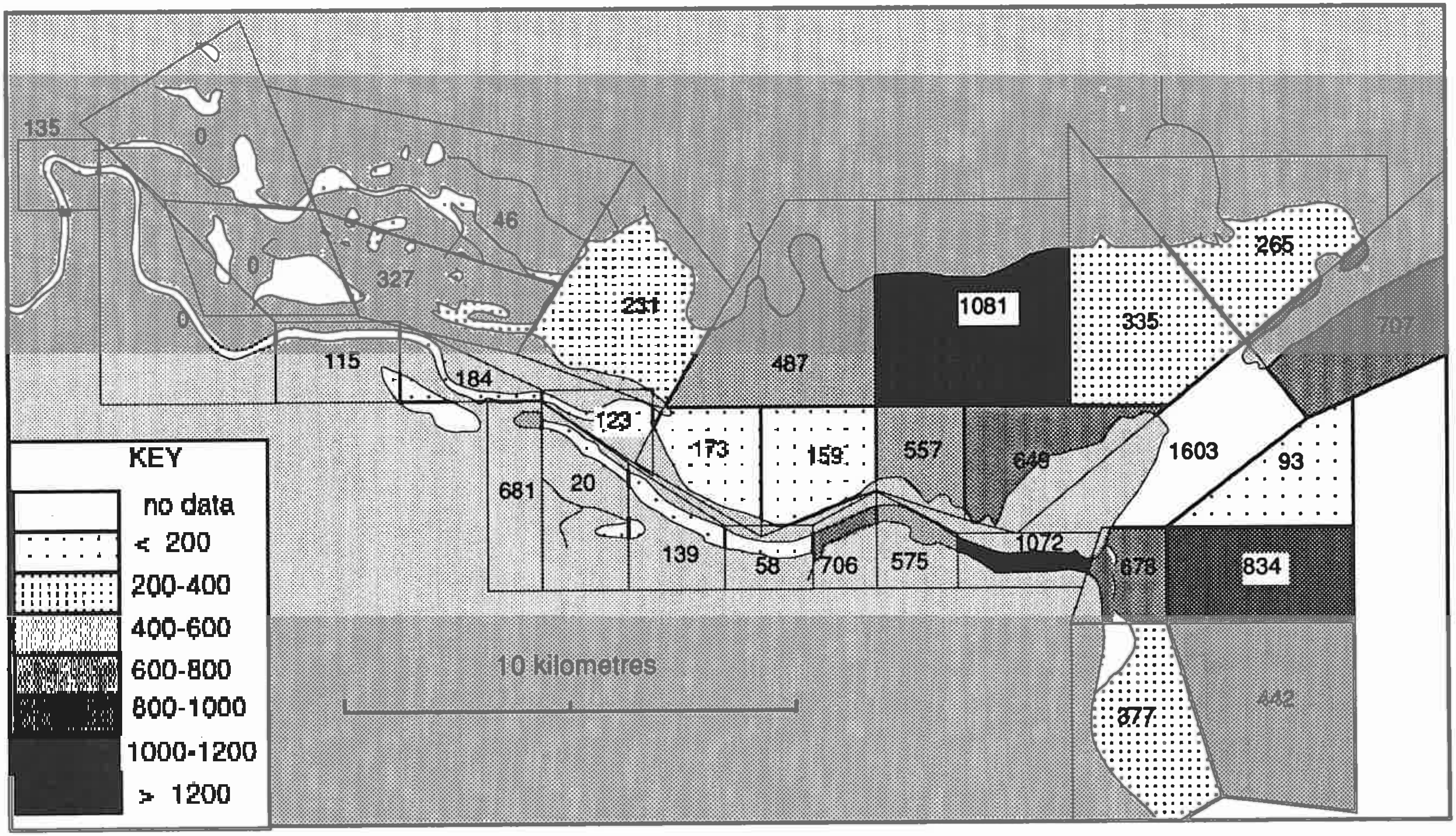


Figure 5-3. Sampling density (total number of measurements) of WQSAL for Nueces system

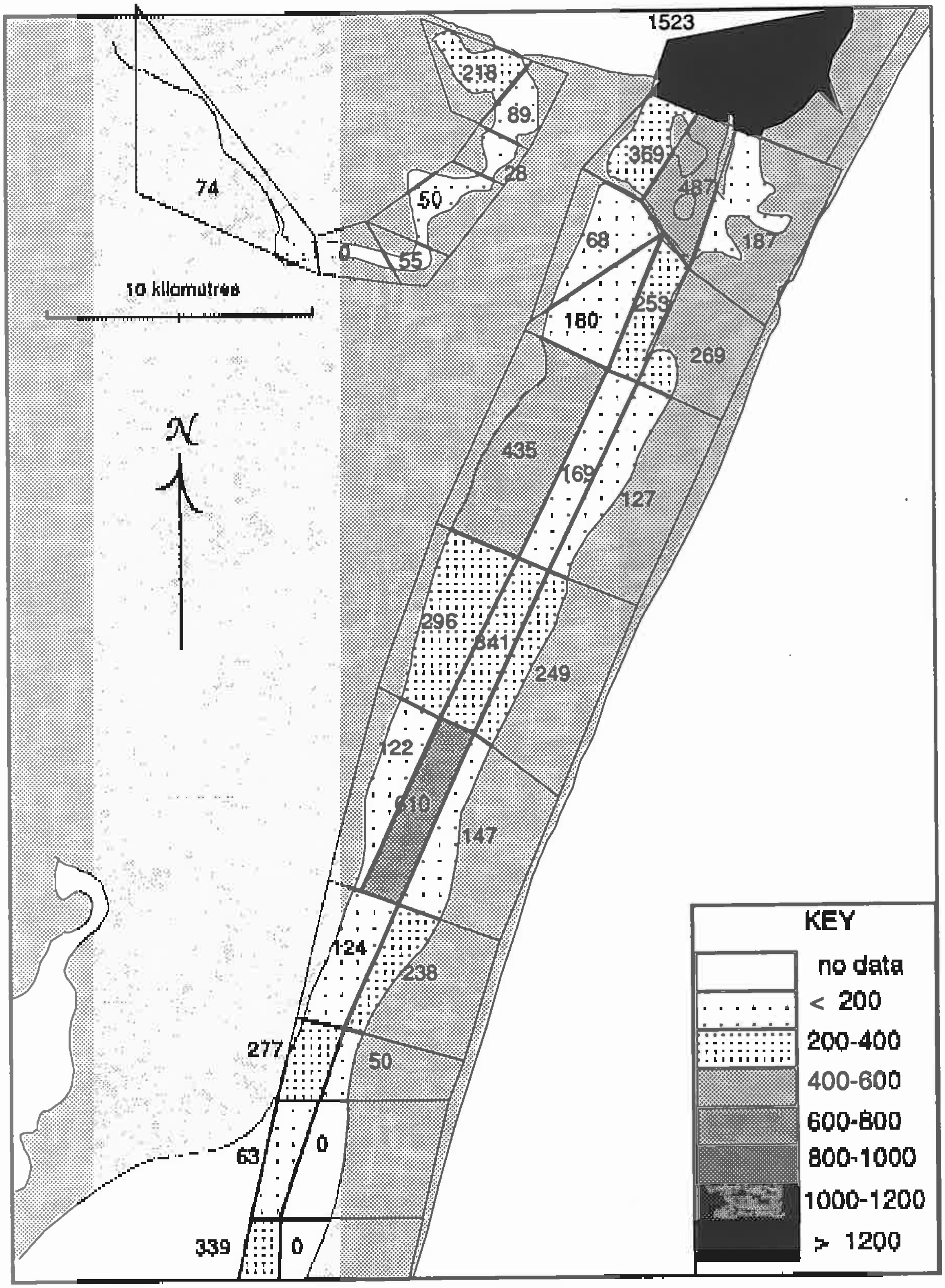


Figure 5-4. Sampling density of WQSAL for Upper Laguna Madre

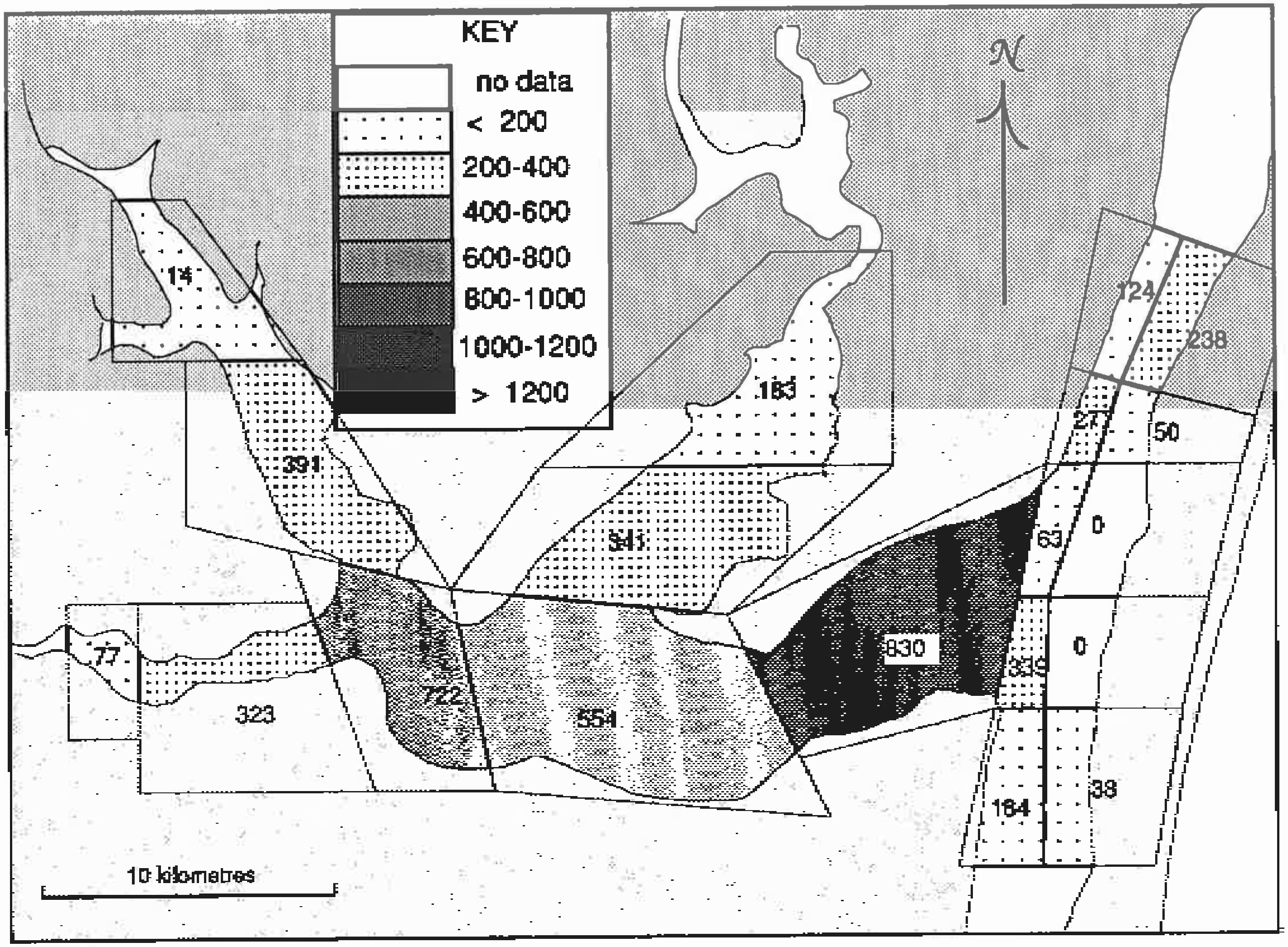


Figure 5-5. Sampling density (total number of measurements) of WQSAL for Baffin Bay system

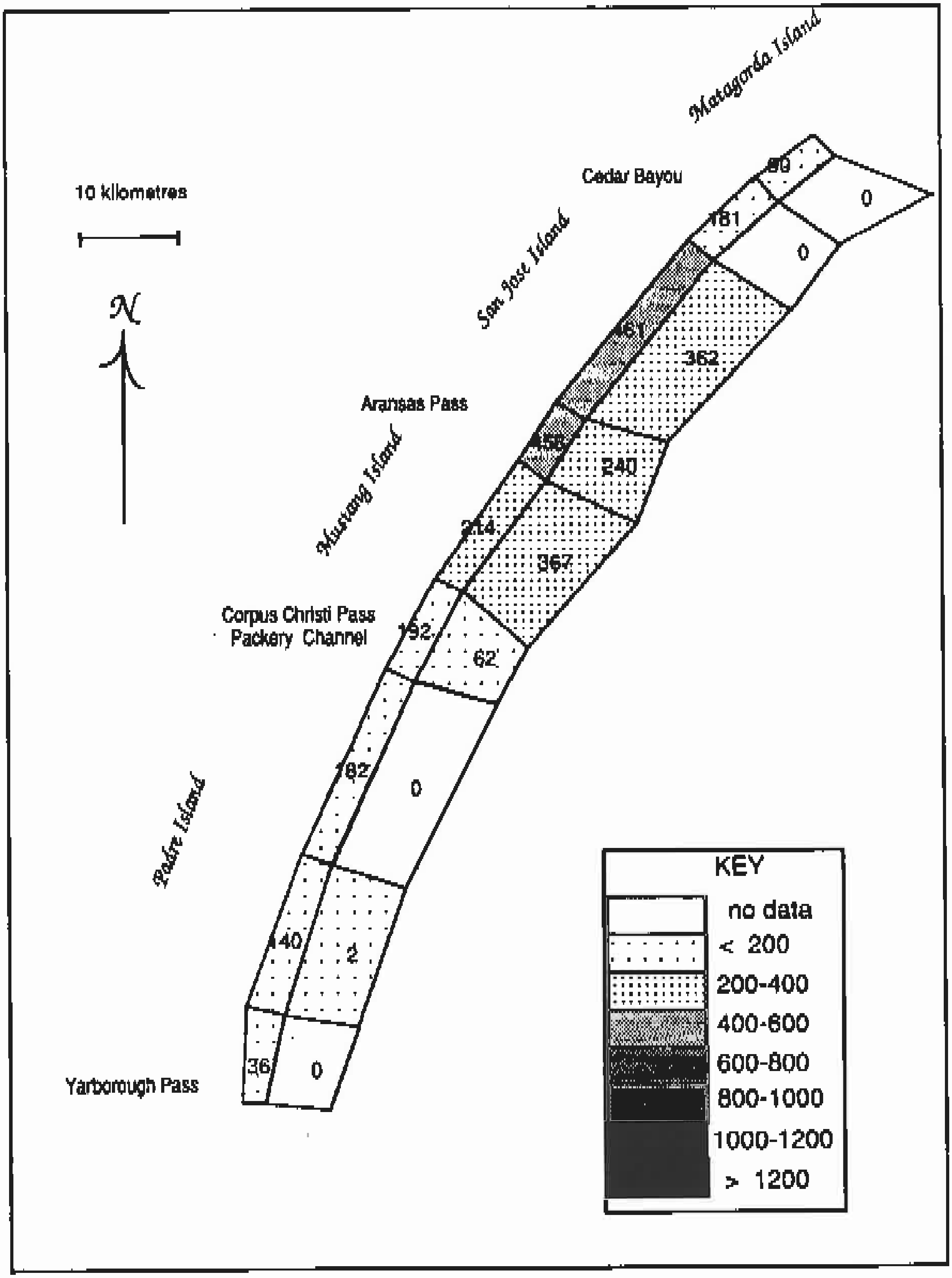


Figure 5-6. Sampling density of WQSAL for Gulf of Mexico

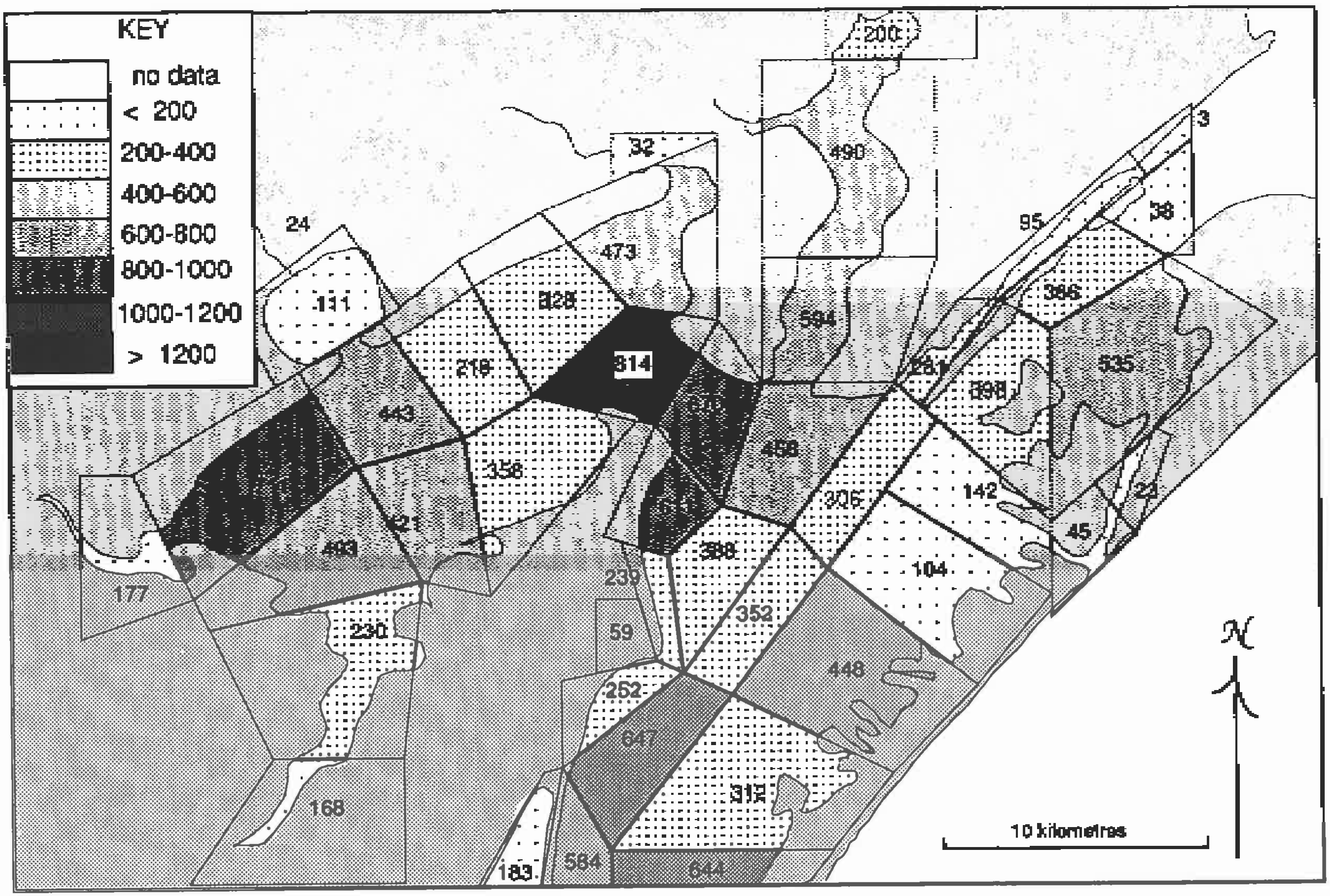


Figure 5-7. Sampling density (total number of measurements) of WQTEMP for Aransas-Copano system

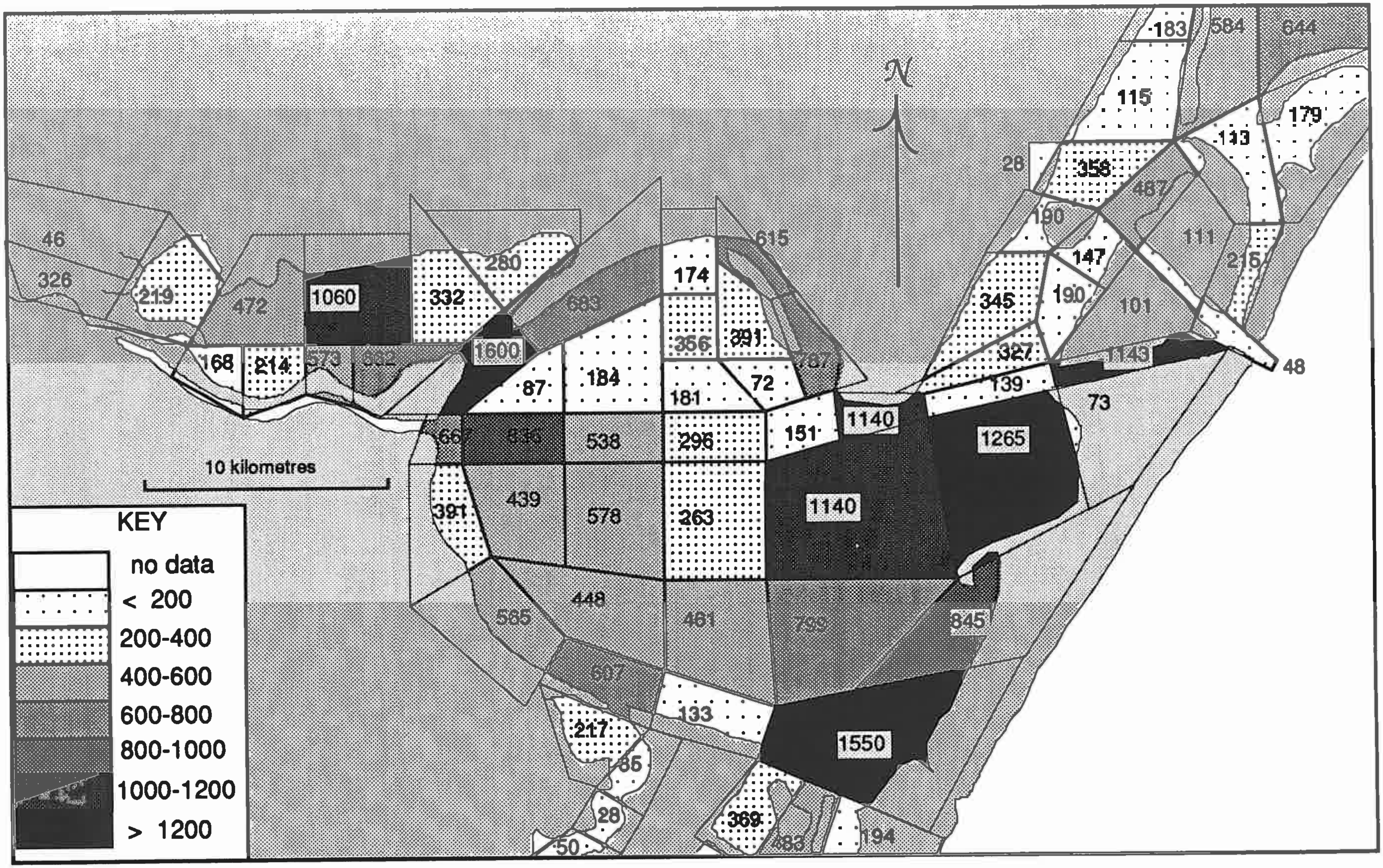


Figure 5-8. Sampling density (total number of measurements) of WQTEMP for Corpus Christi system

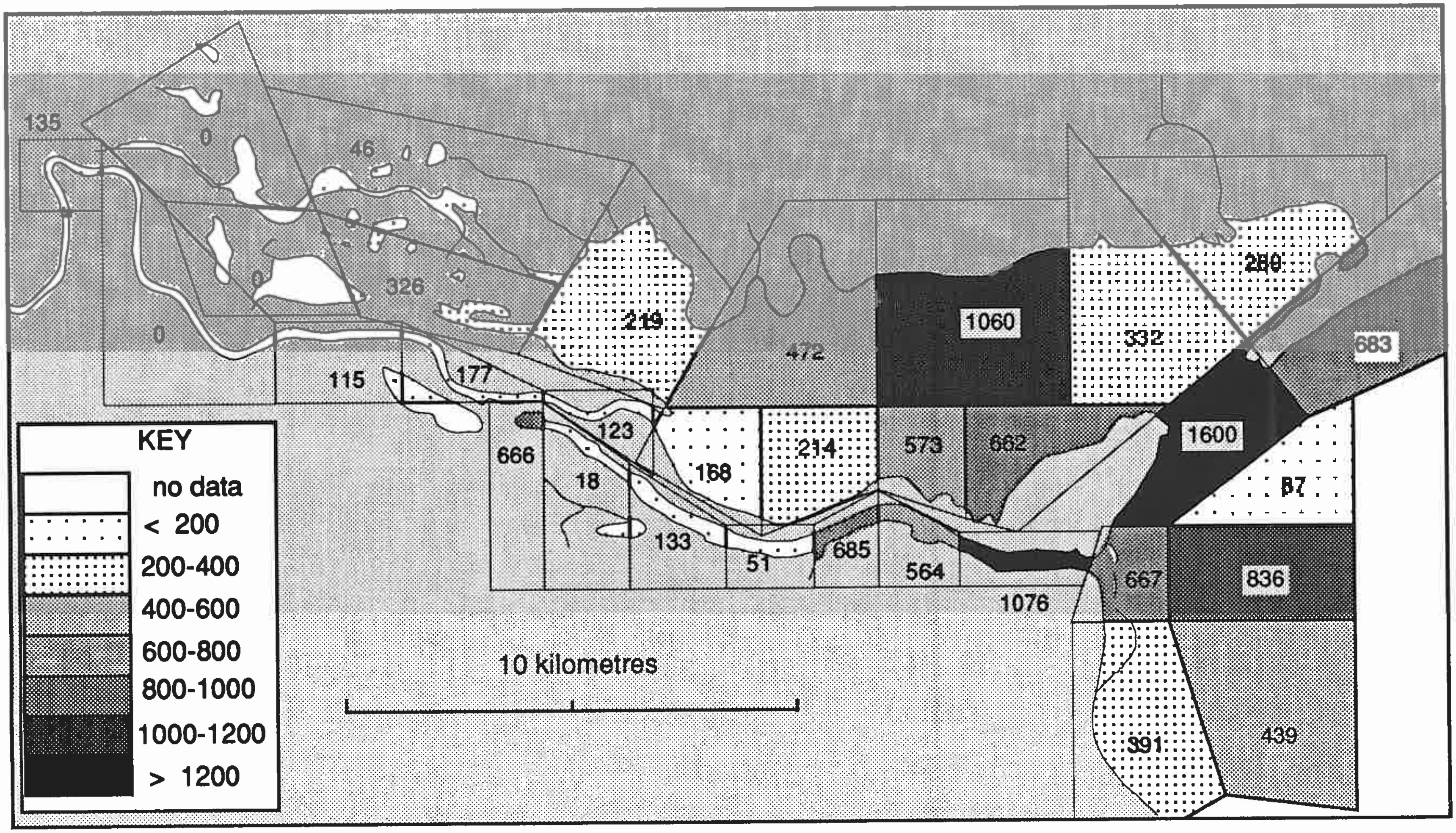


Figure 5-9. Sampling density (total number of measurements) of WQTEMP for Nueces system

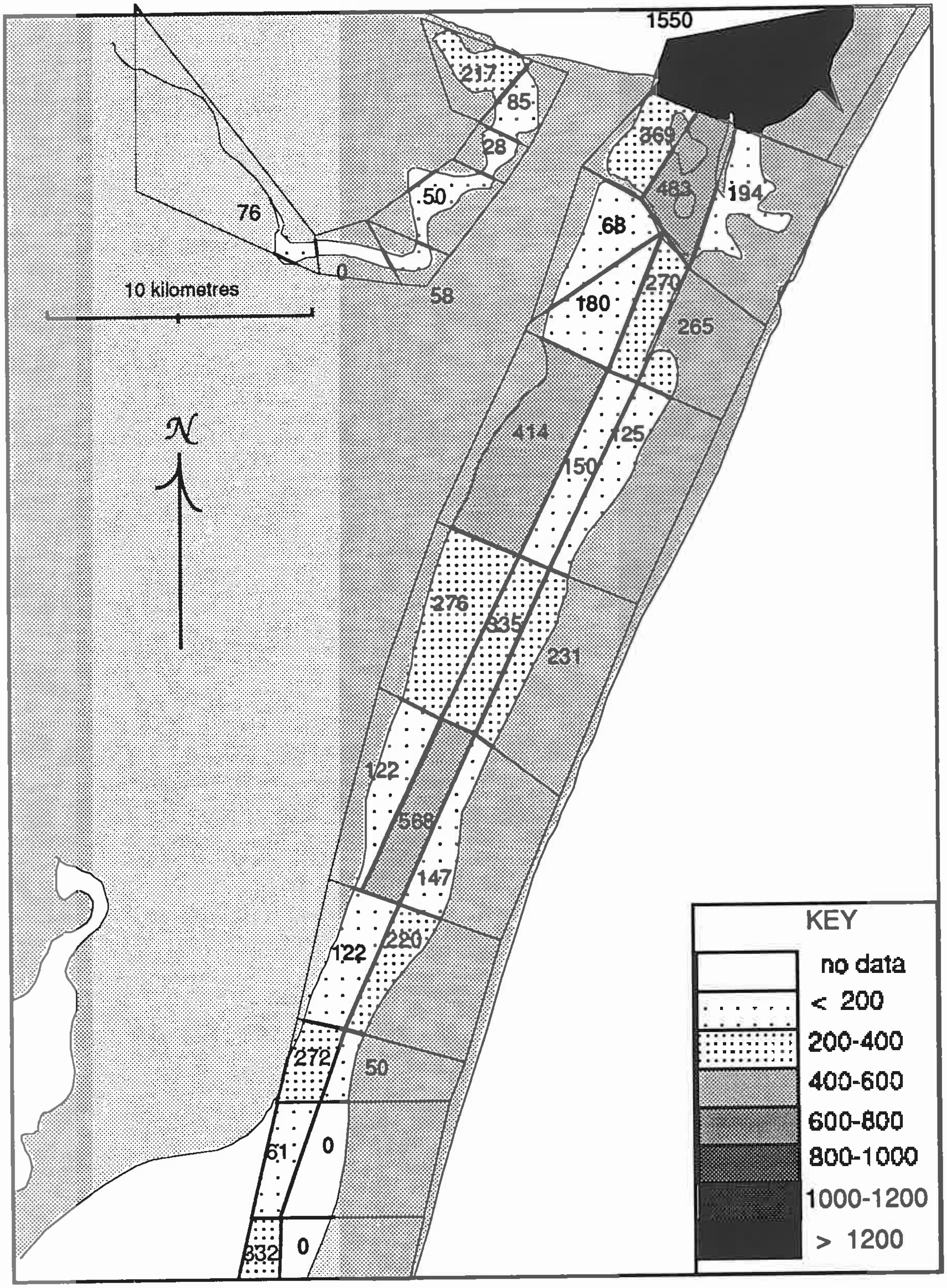


Figure 5-10. Sampling density of WQTEMP for Upper Laguna Madre

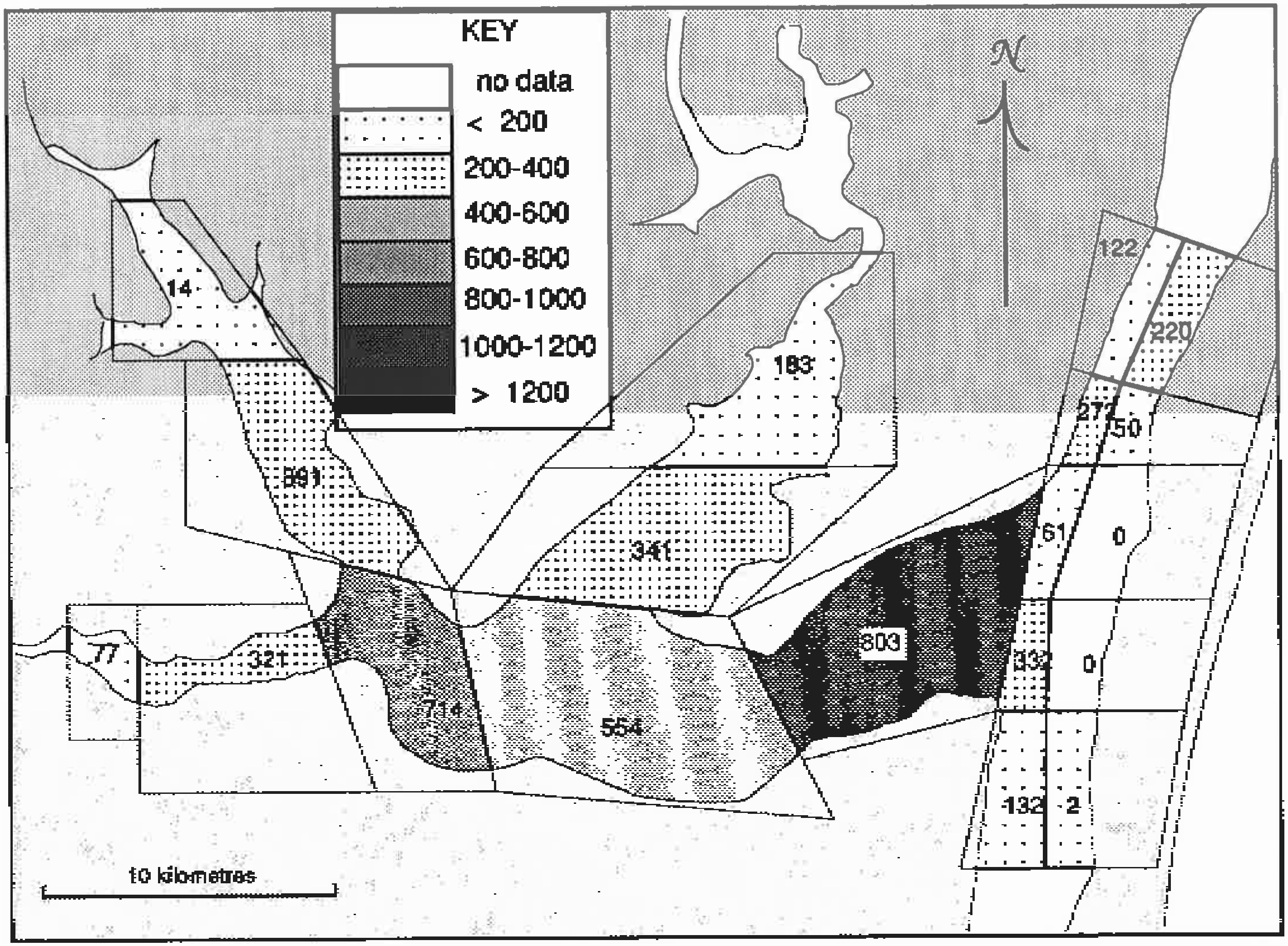


Figure 5-11. Sampling density (total number of measurements) of WQTEMP for Baffin Bay system

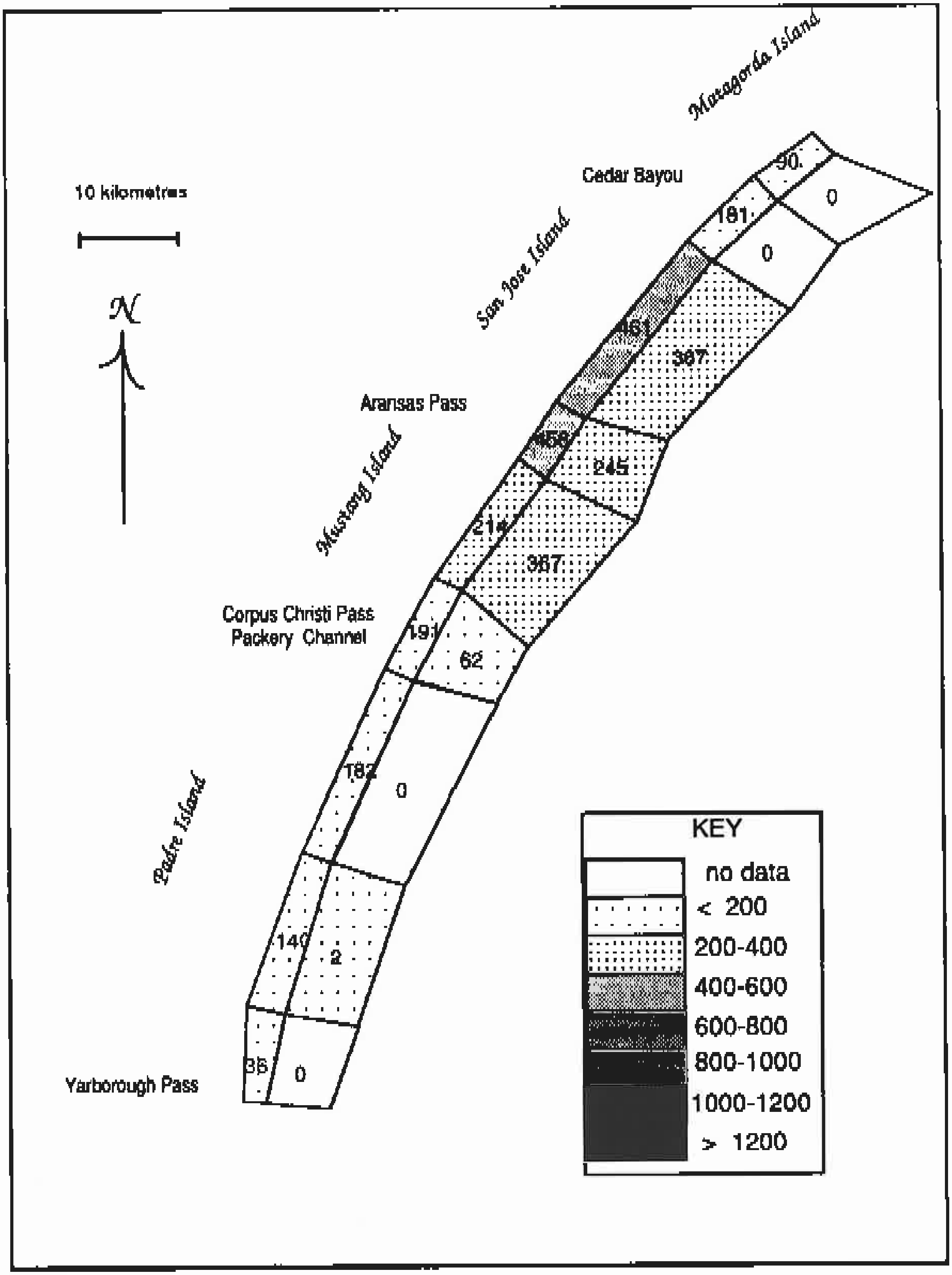


Figure 5-12. Sampling density of WQTEMP for Gulf of Mexico

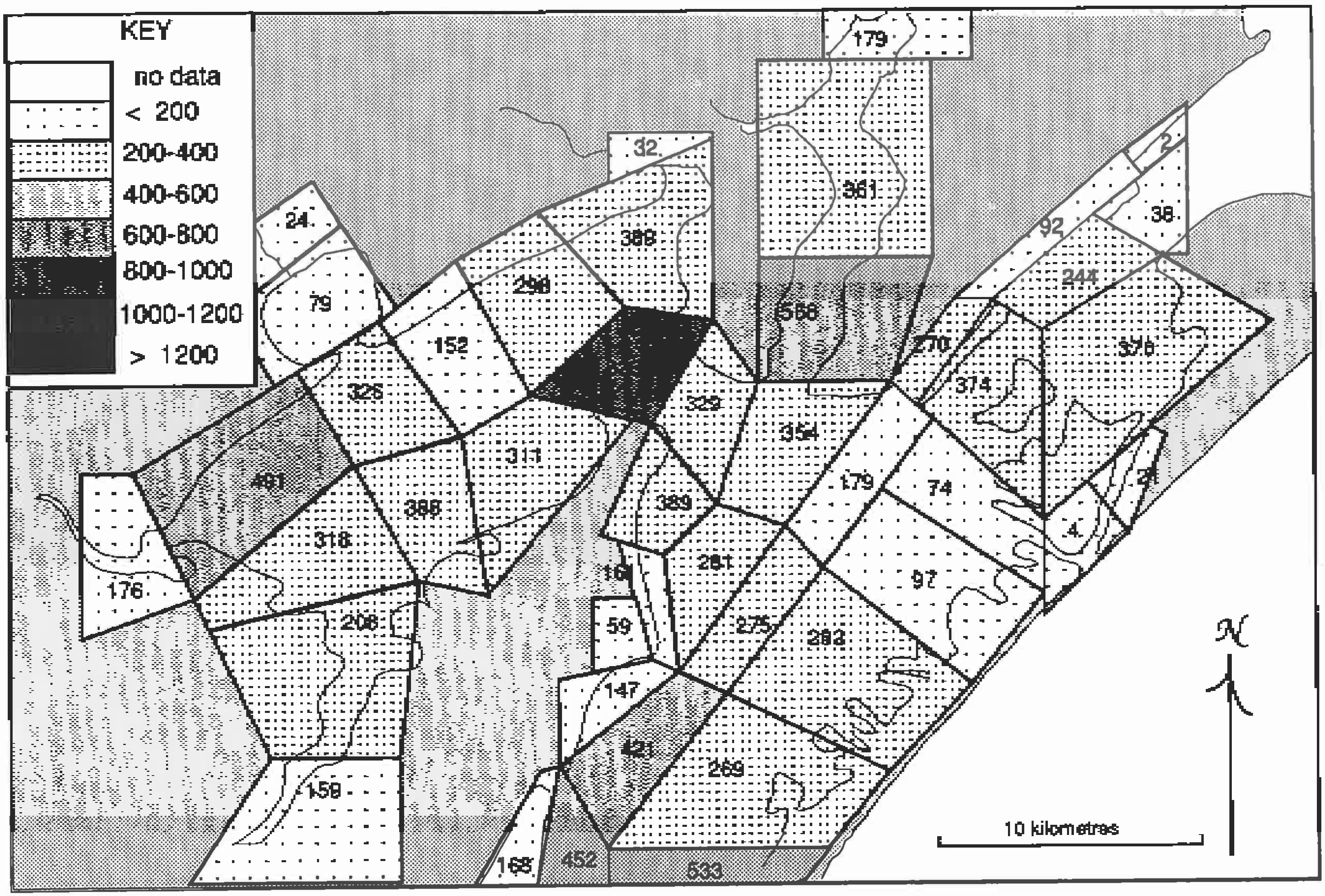


Figure 5-13. Sampling density of WQDO for Aransas-Copano system

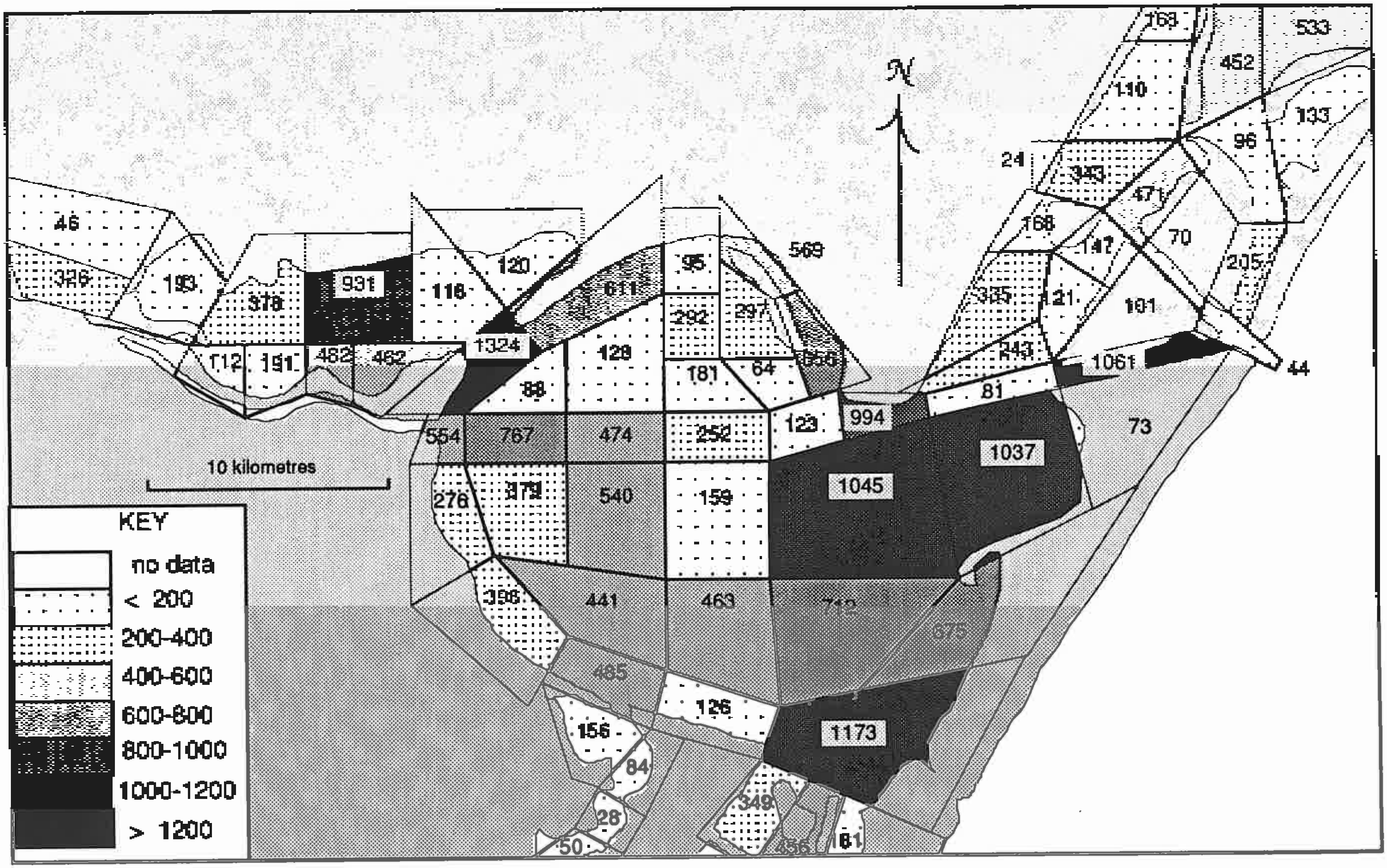


Figure 5-14. Sampling density of WQDO for Corpus Christi system

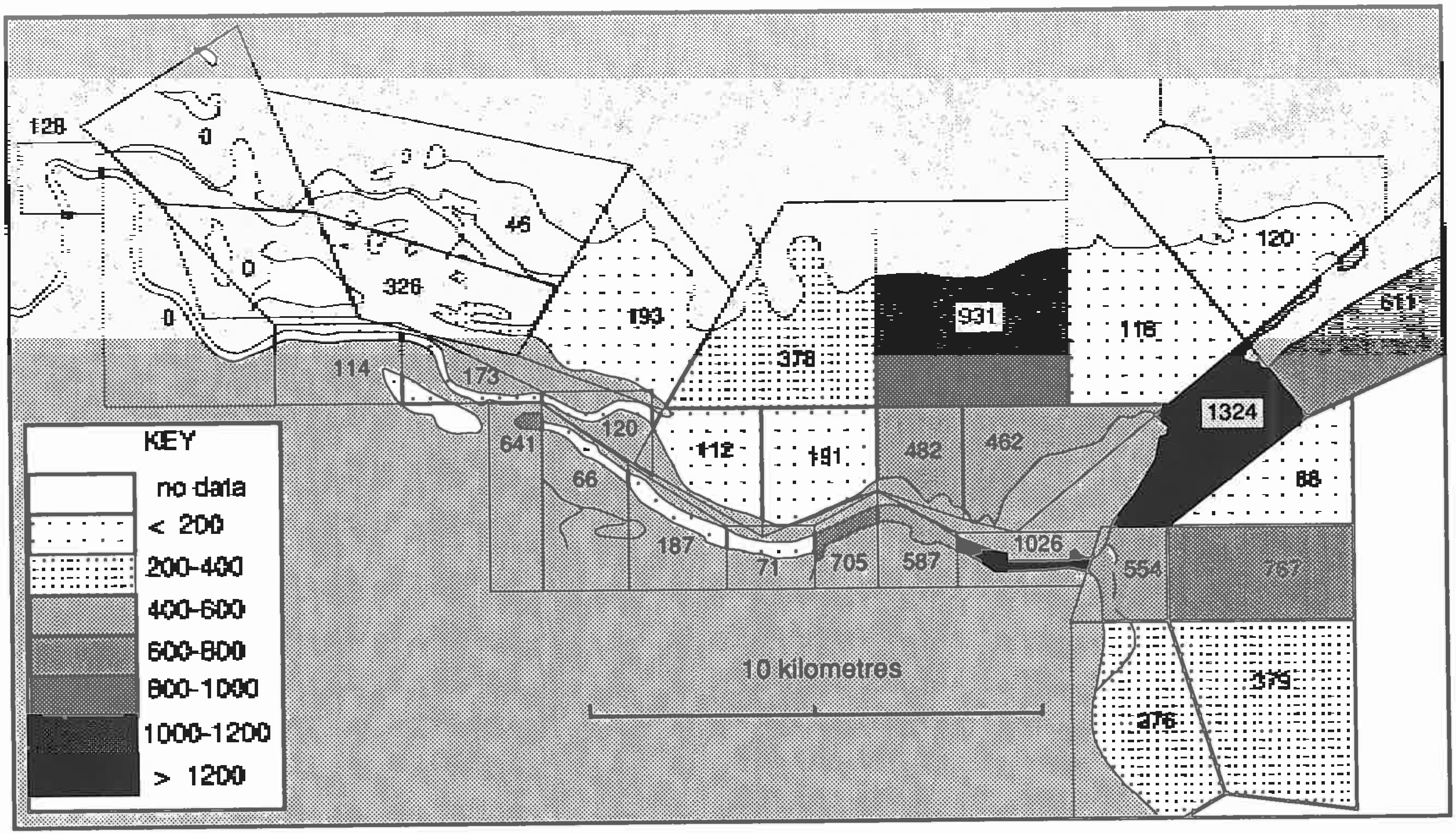


Figure 5-15. Sampling density of WQDO for Nueces Bay region, including Inner Harbor

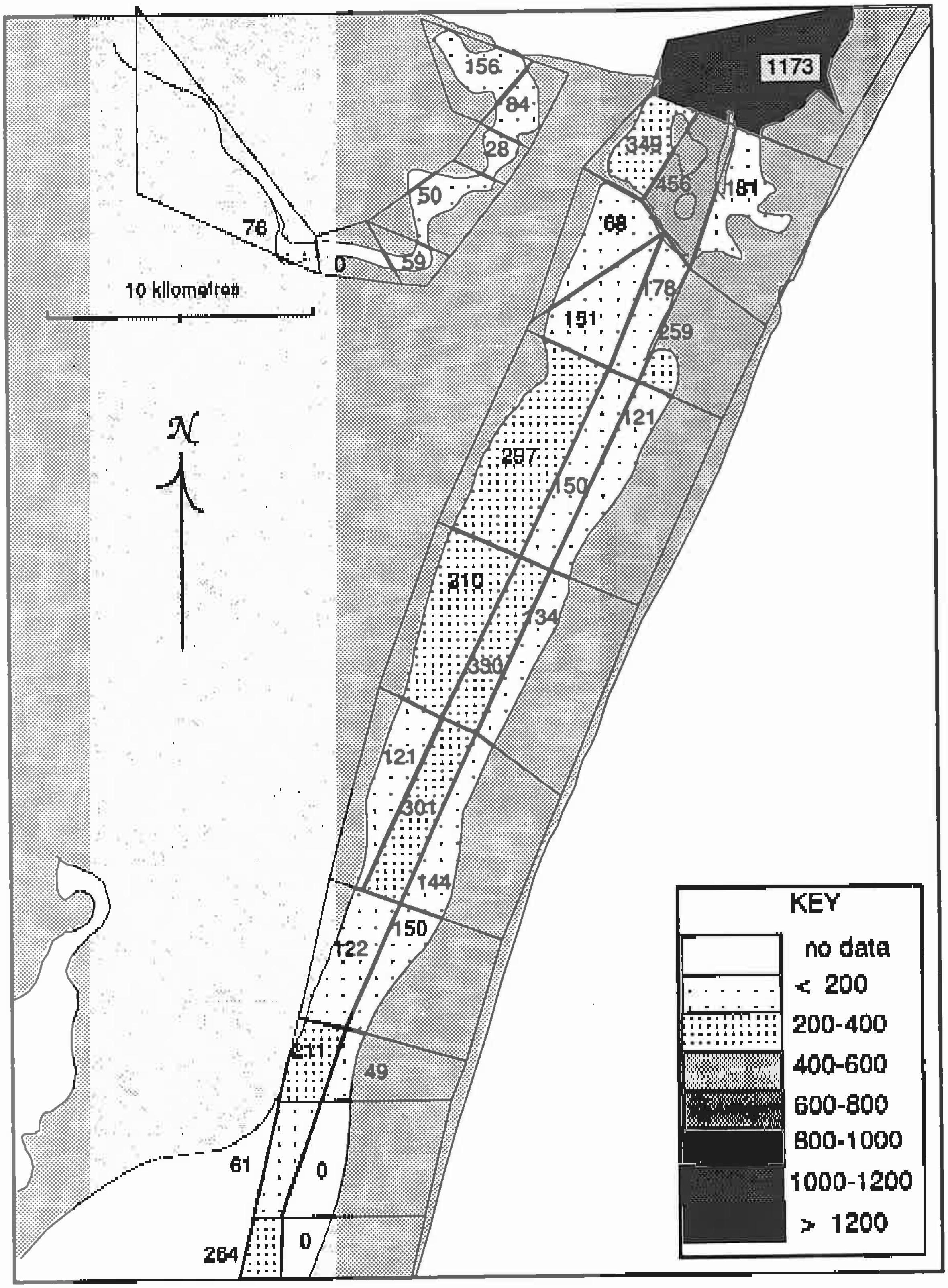


Figure 5-16. Sampling density of WQDO for Upper Laguna Madre and Oso Bay

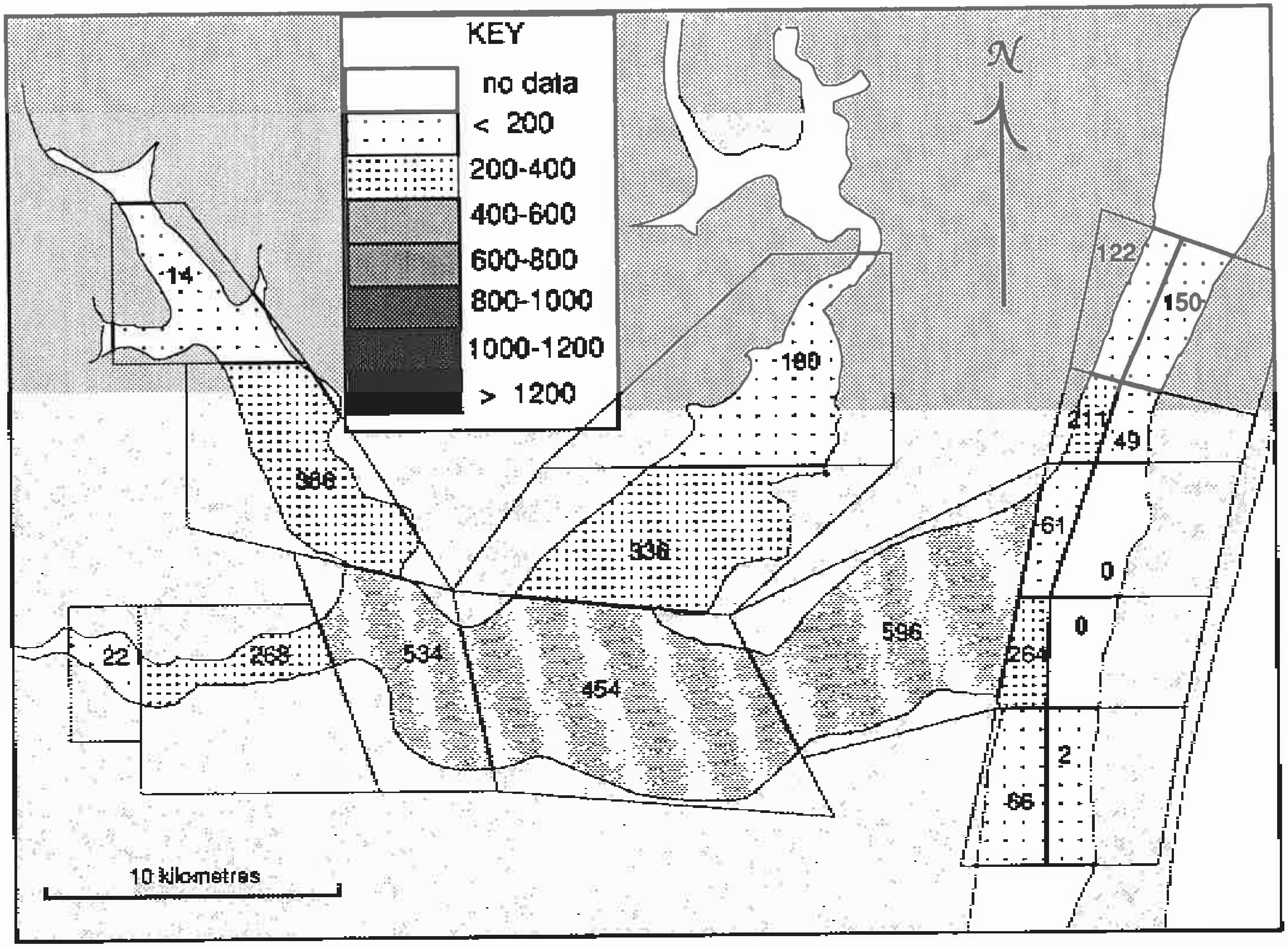


Figure 5-17. Sampling density of WQDO for Baffin Bay region

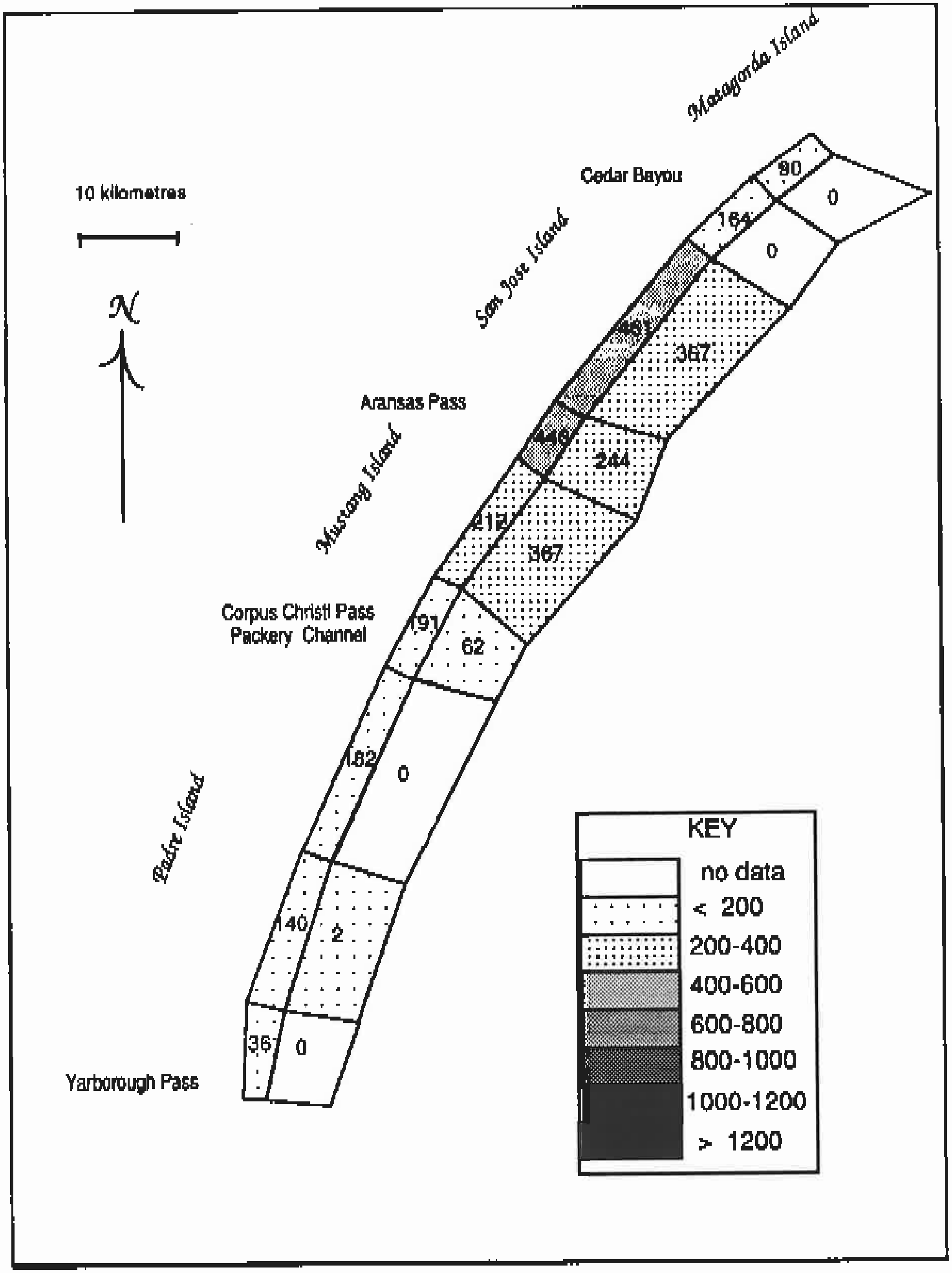


Figure 5-18. Sampling density of WQDO for Gulf of Mexico

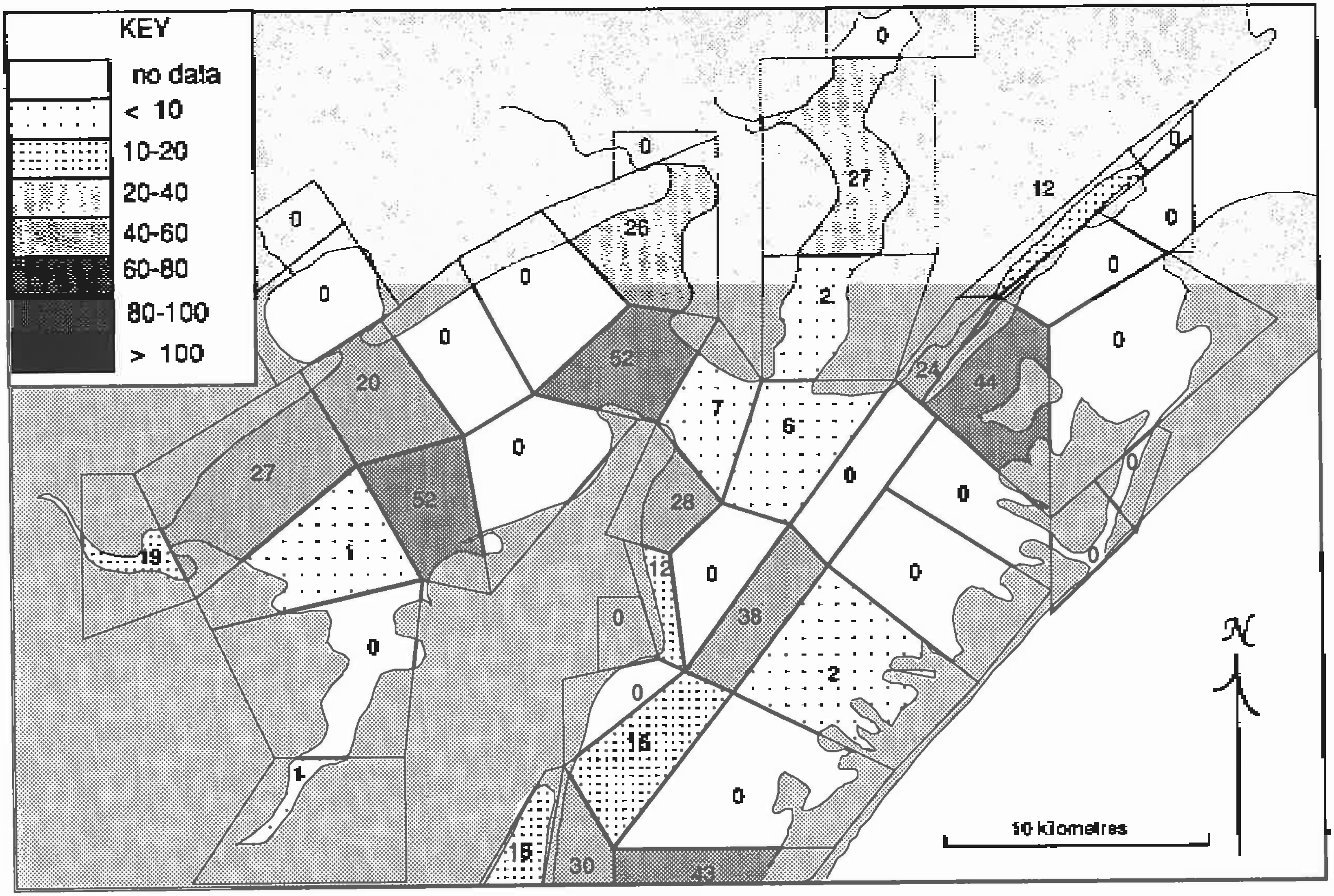


Figure 5-19. Sampling density of WQBOD5 for Aransas-Copano system

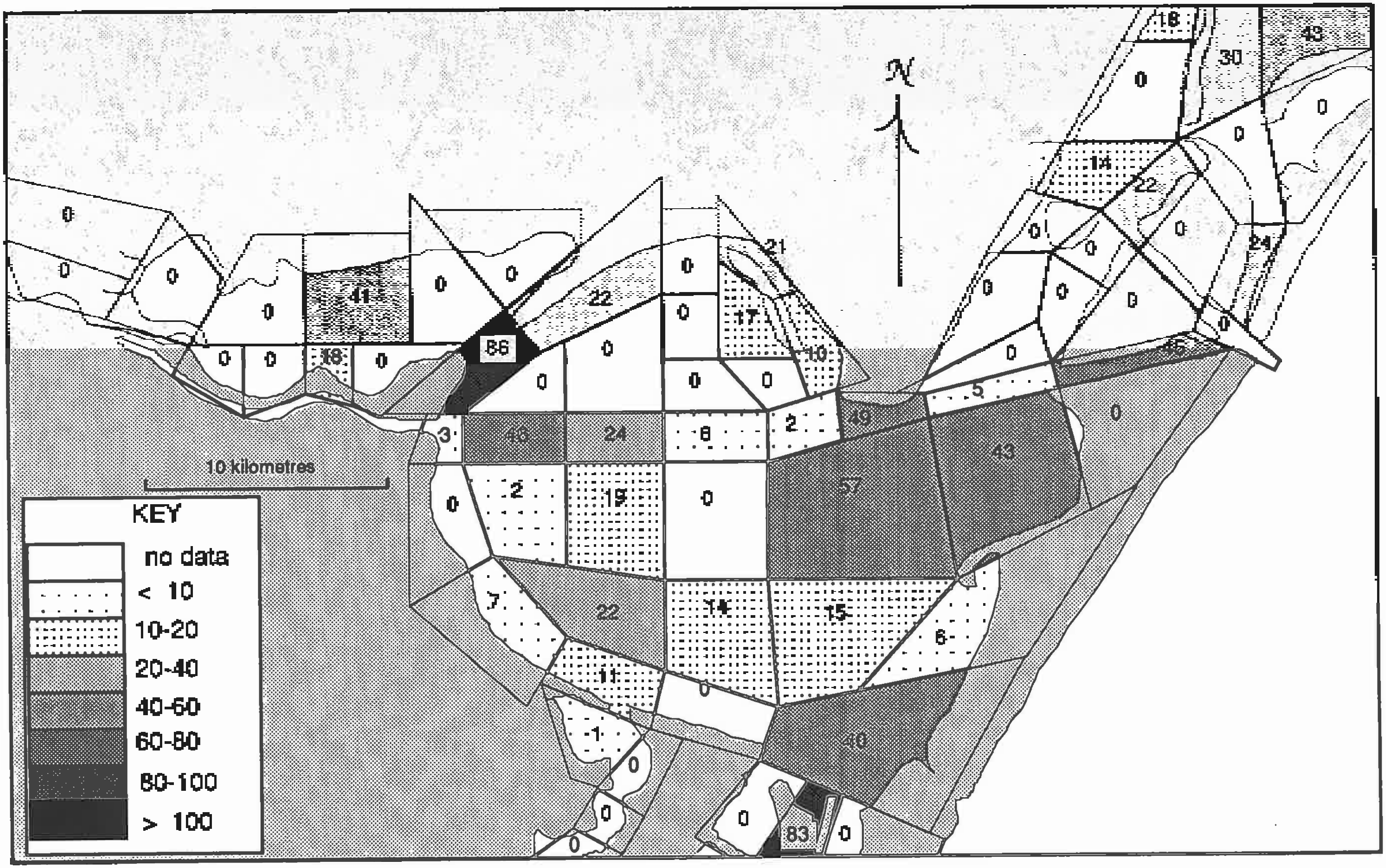


Figure 5-20. Sampling density of WQBOD5 for Corpus Christi system

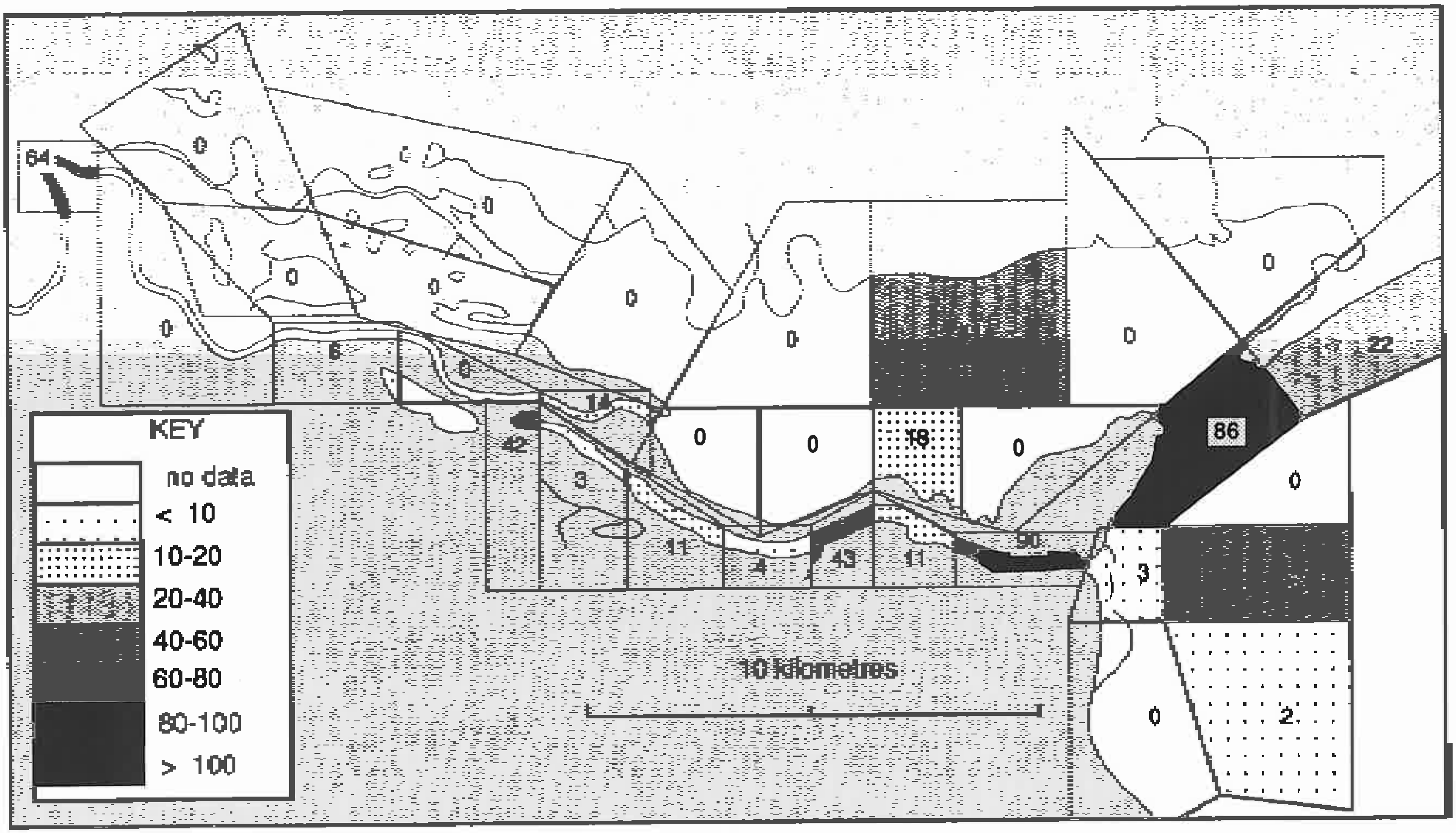


Figure 5-21. Sampling density of WQBOD5 for Nueces Bay region, including Inner Harbor

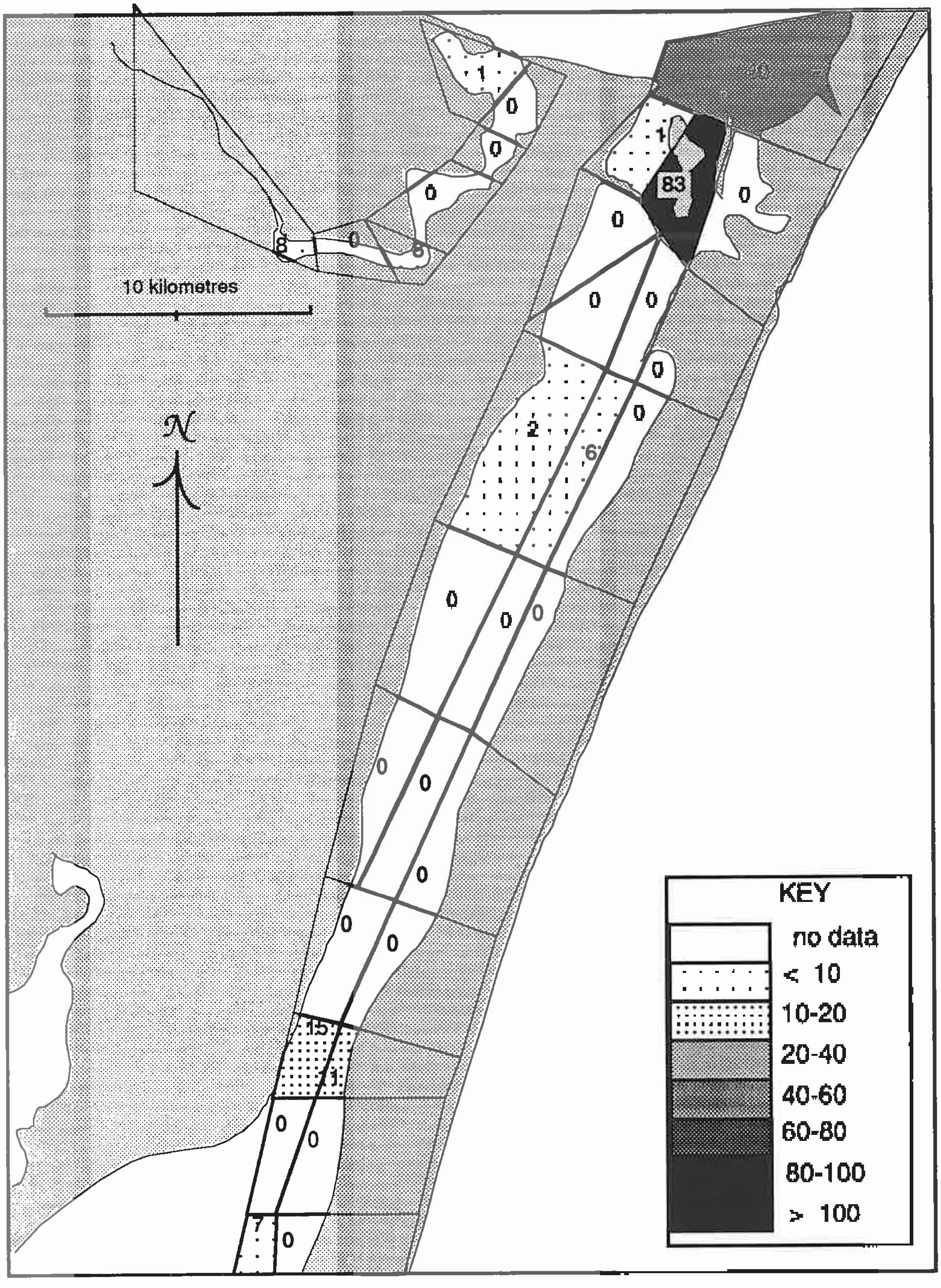


Figure 5-22. Sampling density of WQBOD5 for Upper Laguna Madre and Oso Bay

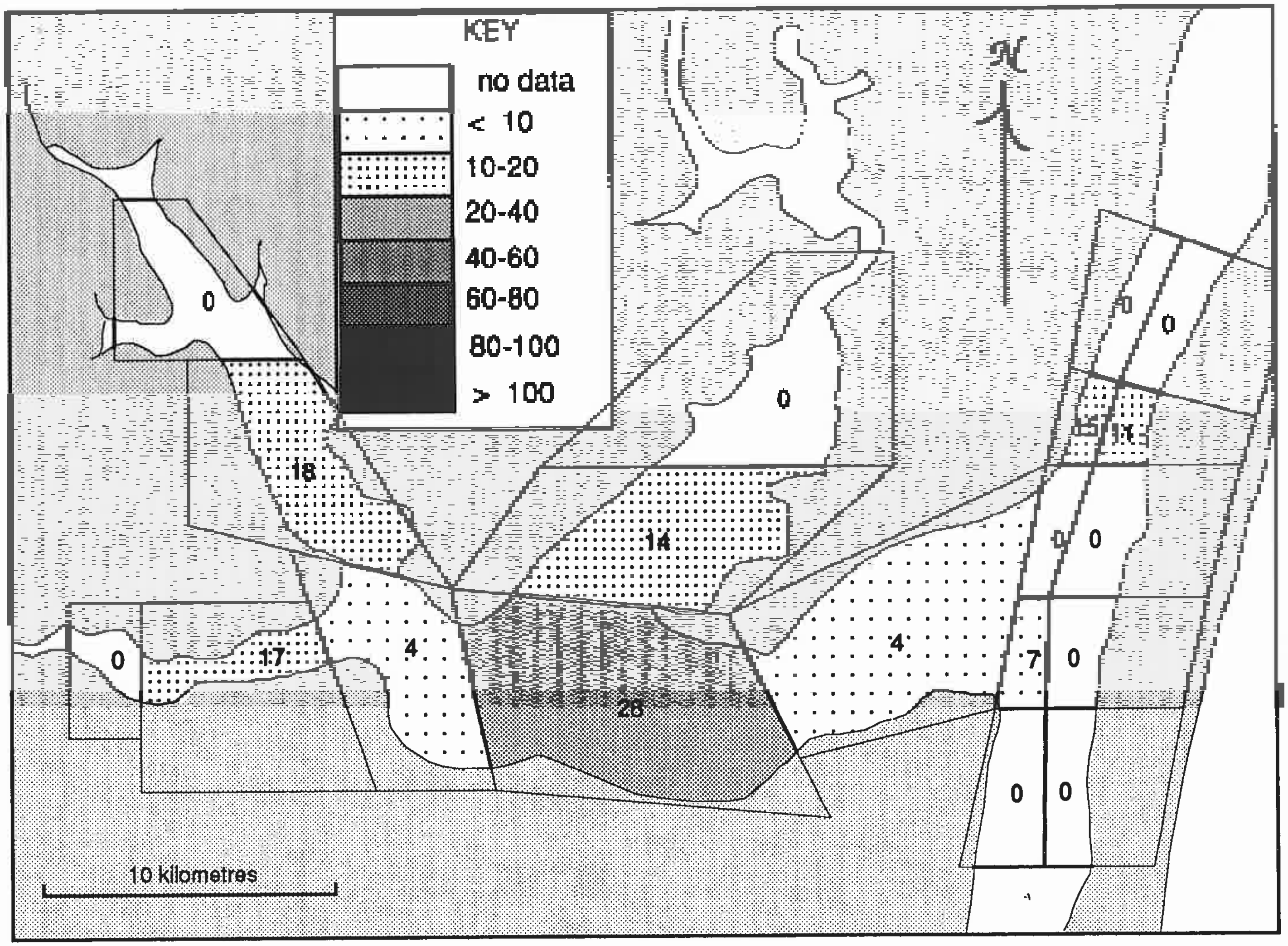


Figure 5-23. Sampling density of WQBOD5 for Baffin Bay region

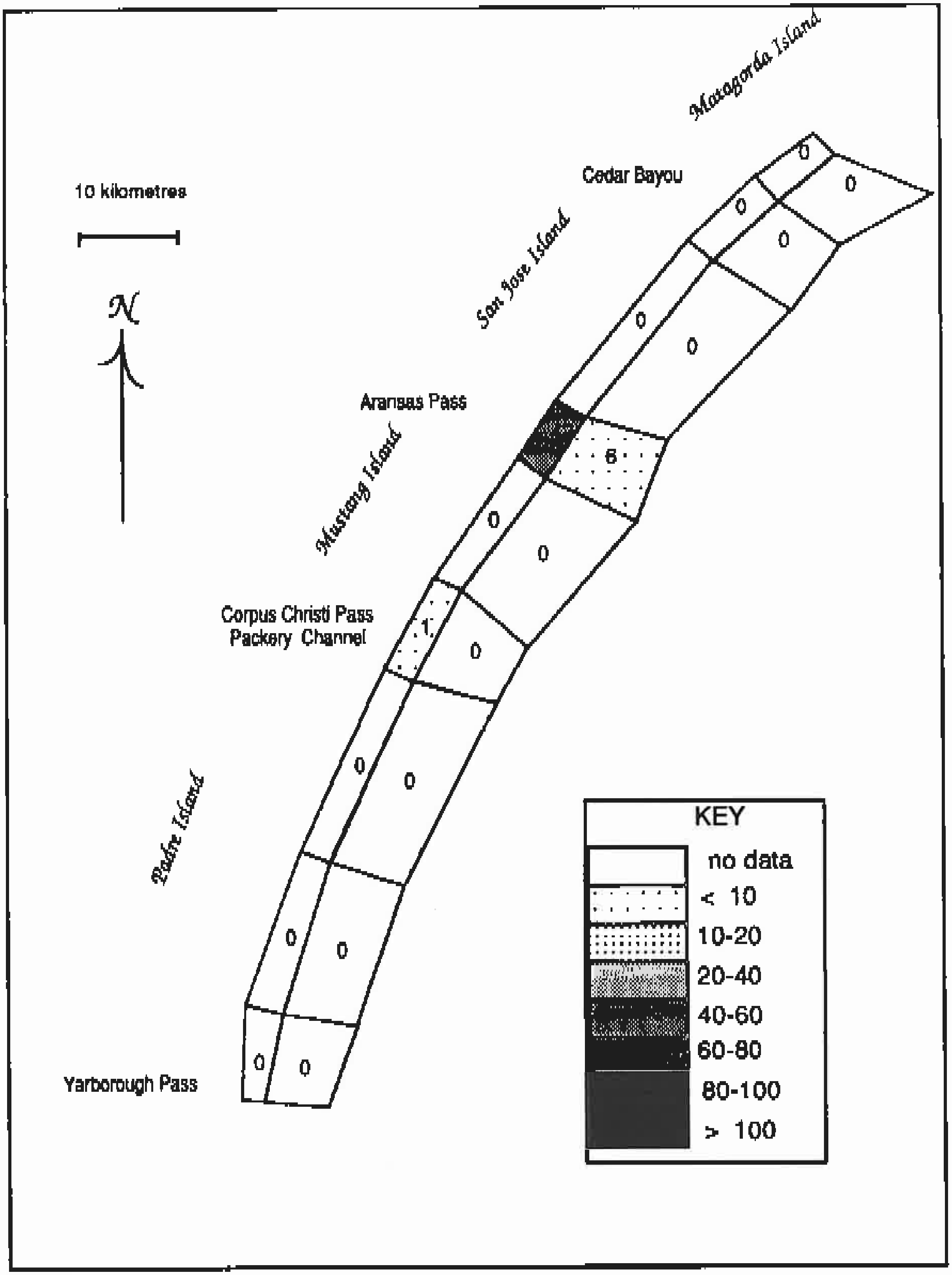


Figure 5-24. Sampling density of WQBOD5 for Gulf of Mexico

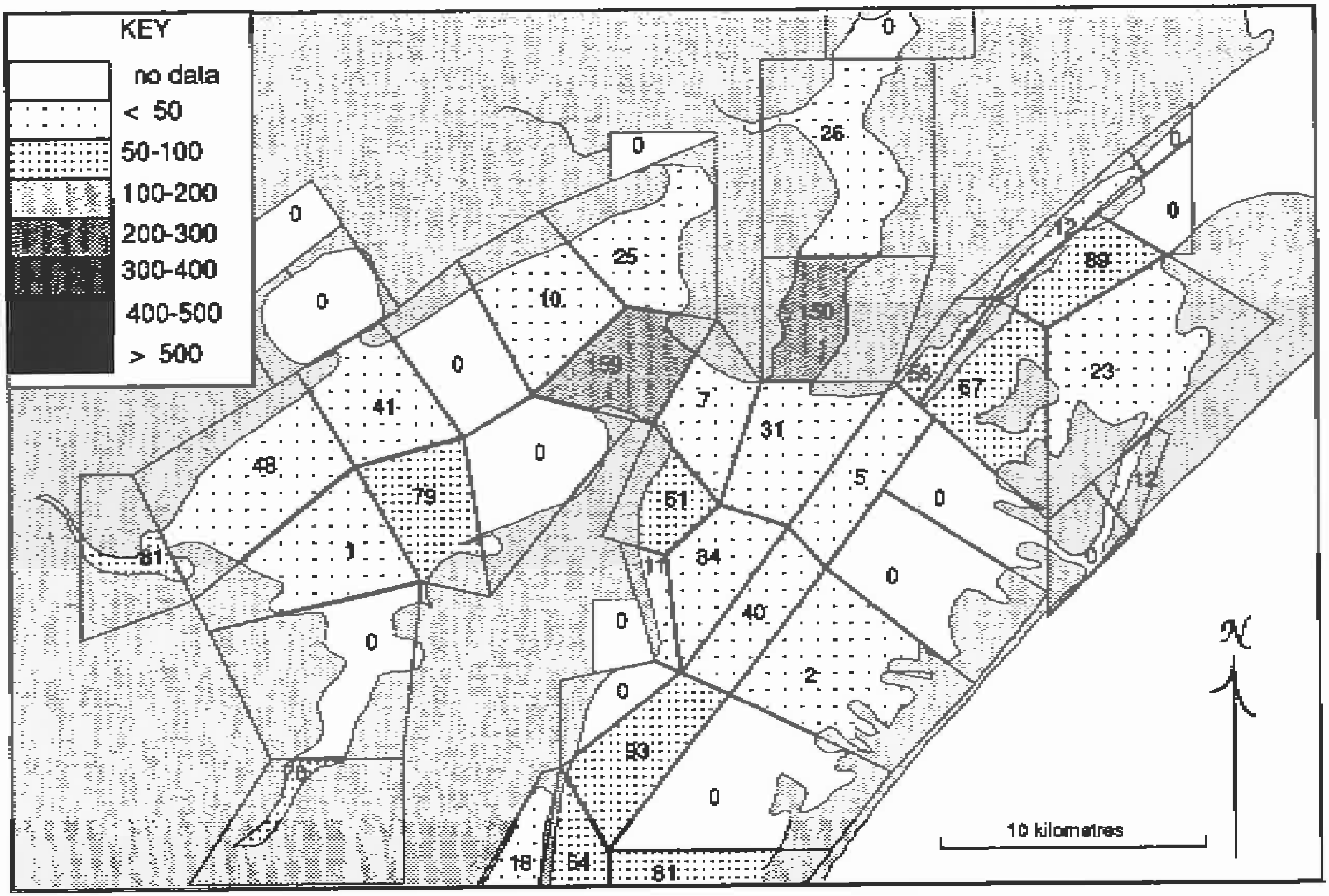


Figure 5-25. Sampling density of WQAMMN for Aransas-Copano system

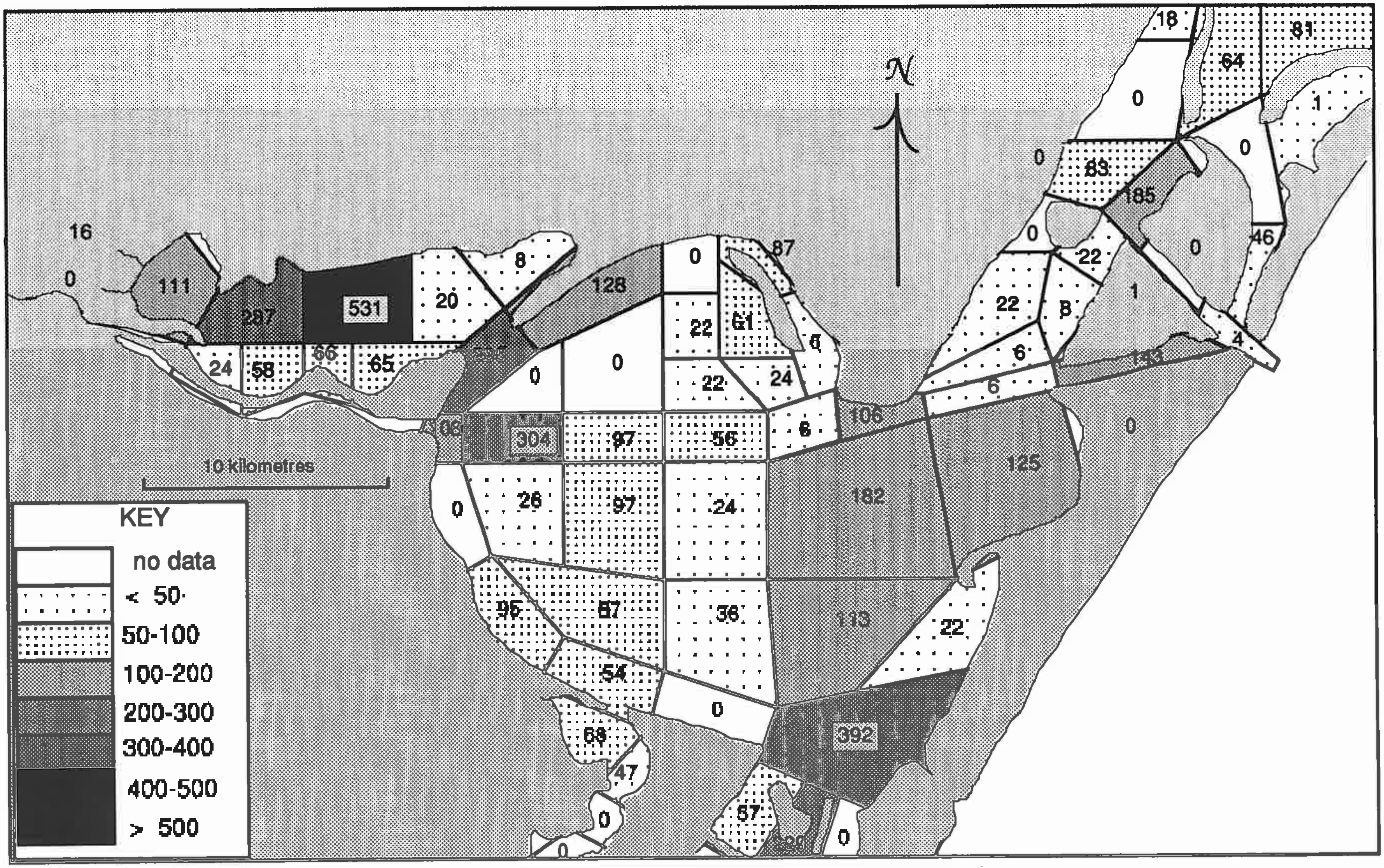


Figure 5-26. Sampling density of WQAMMN for Corpus Christi system

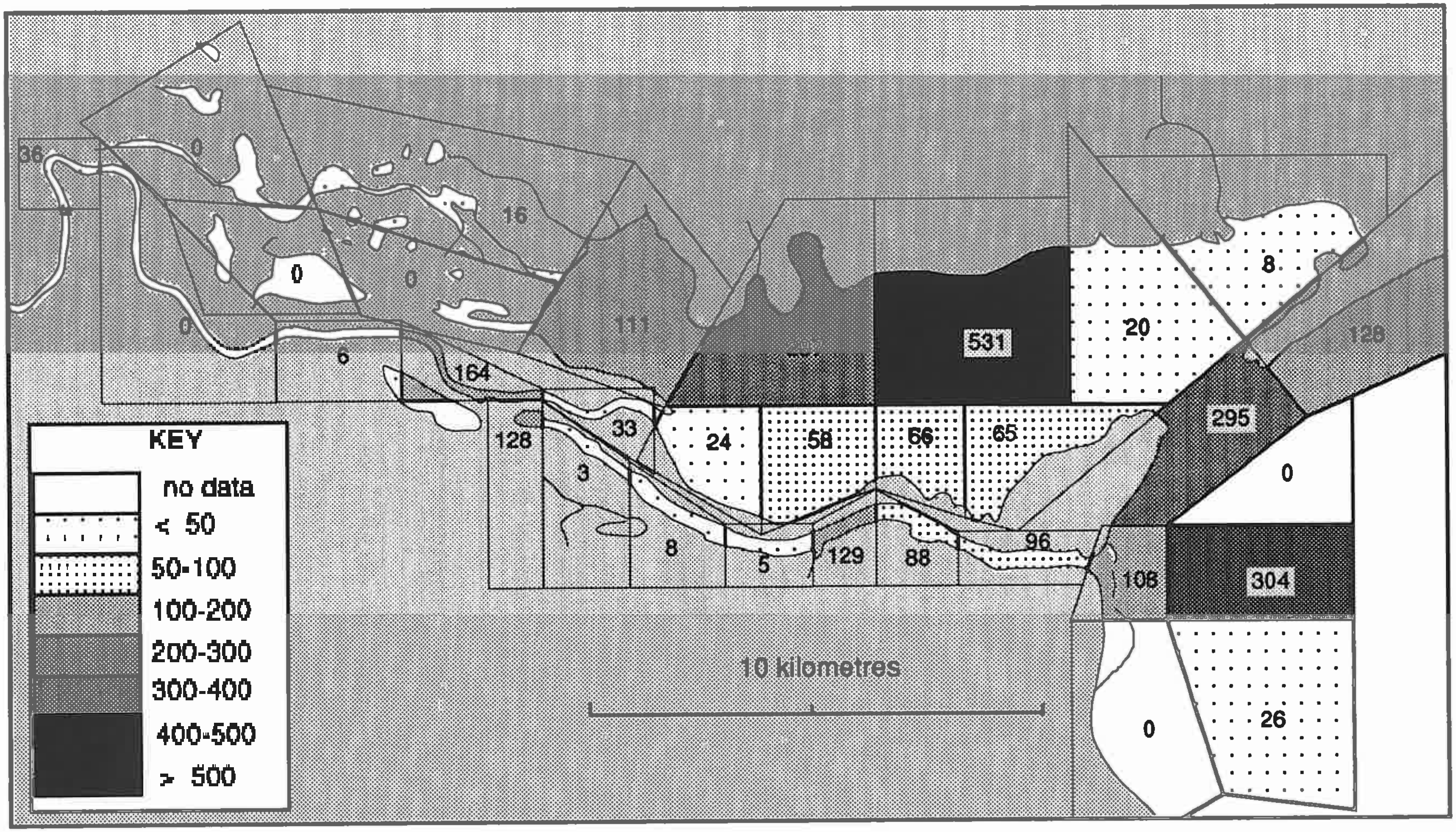


Figure 5-27. Sampling density of WQAMMN for Nueces Bay region, including Inner Harbor

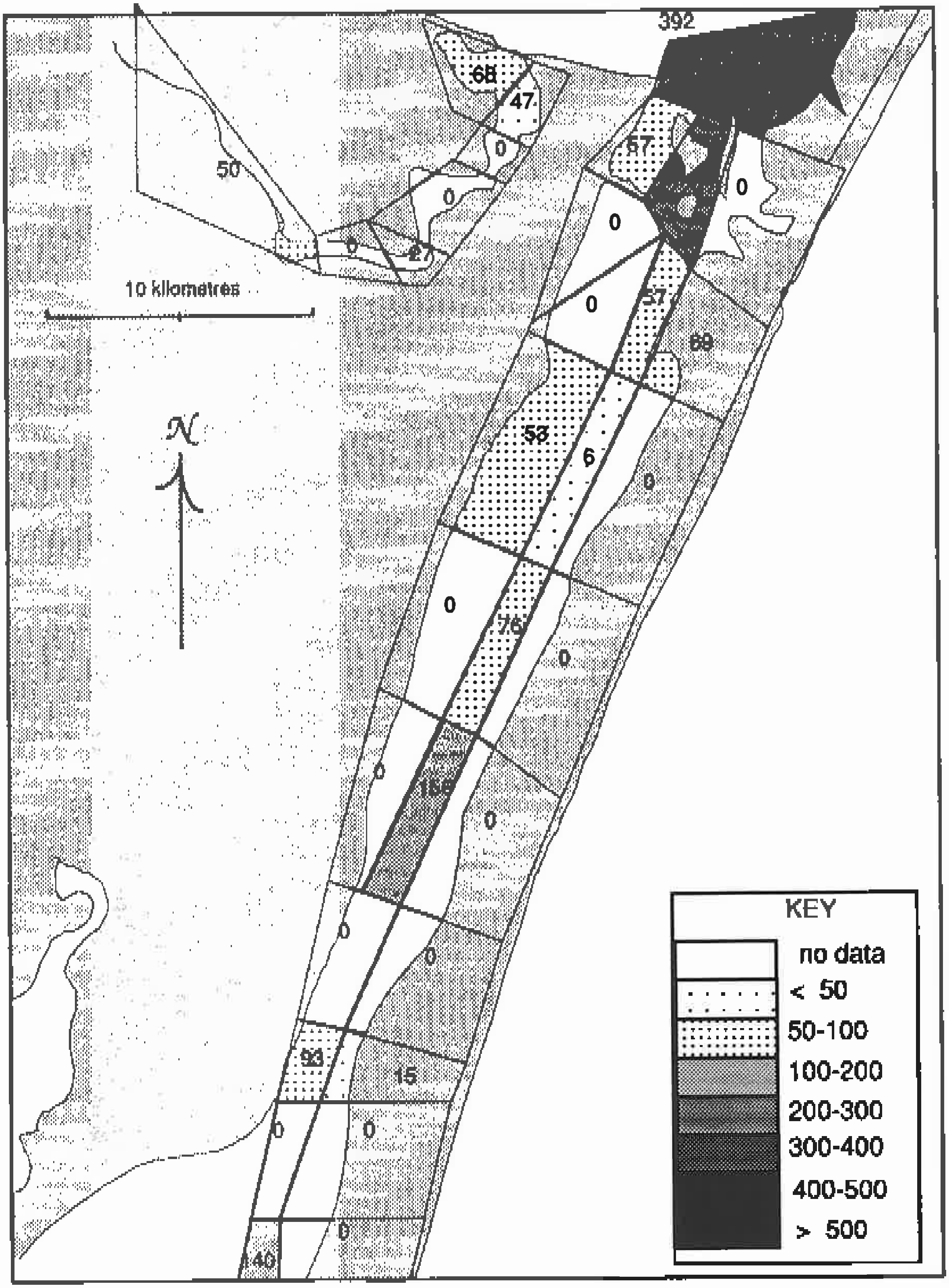


Figure 5-28. Sampling density of WQAMMN for Upper Laguna Madre and Oso Bay

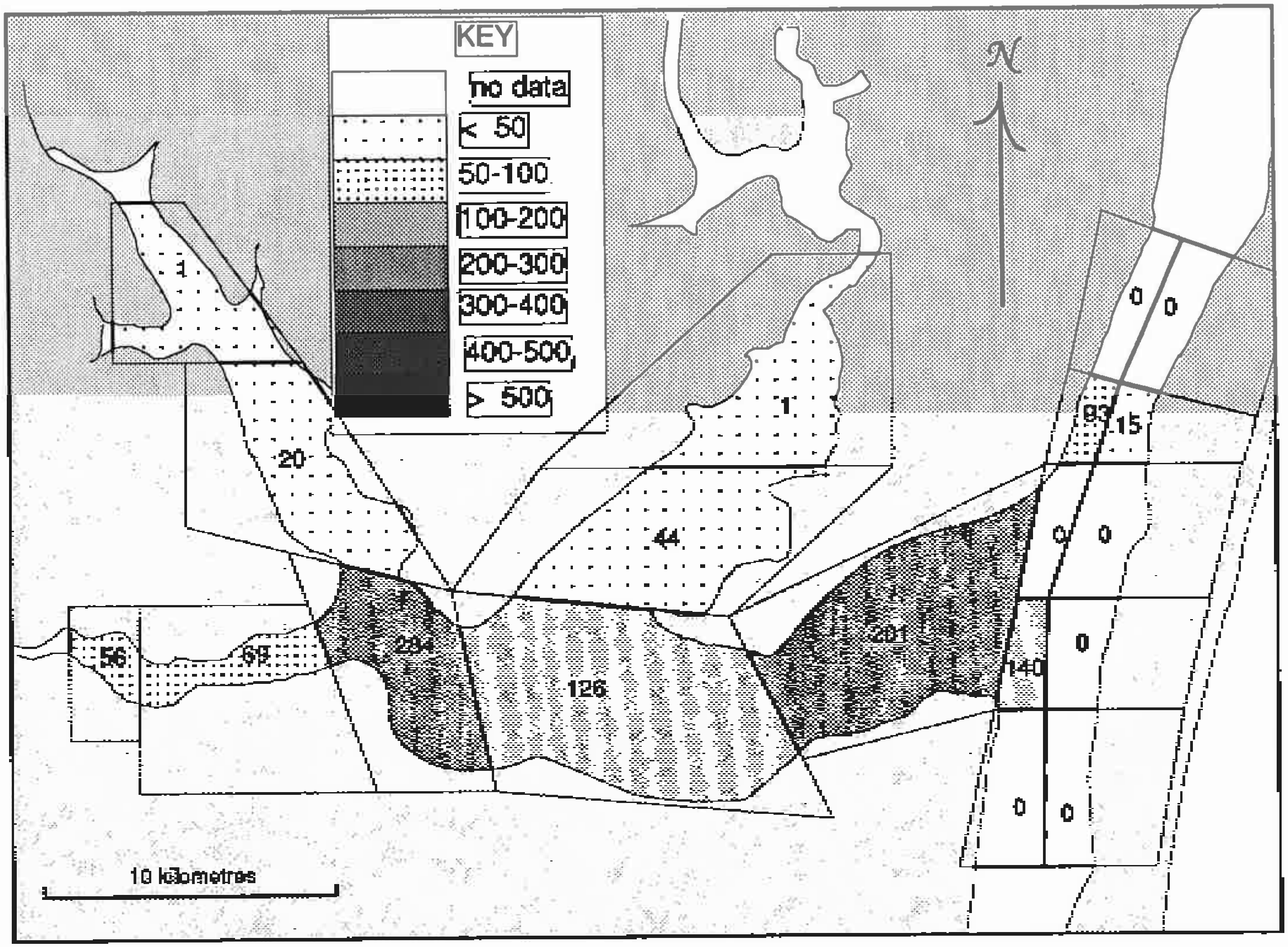


Figure 5-29. Sampling density of WQAMMN for Baffin Bay region

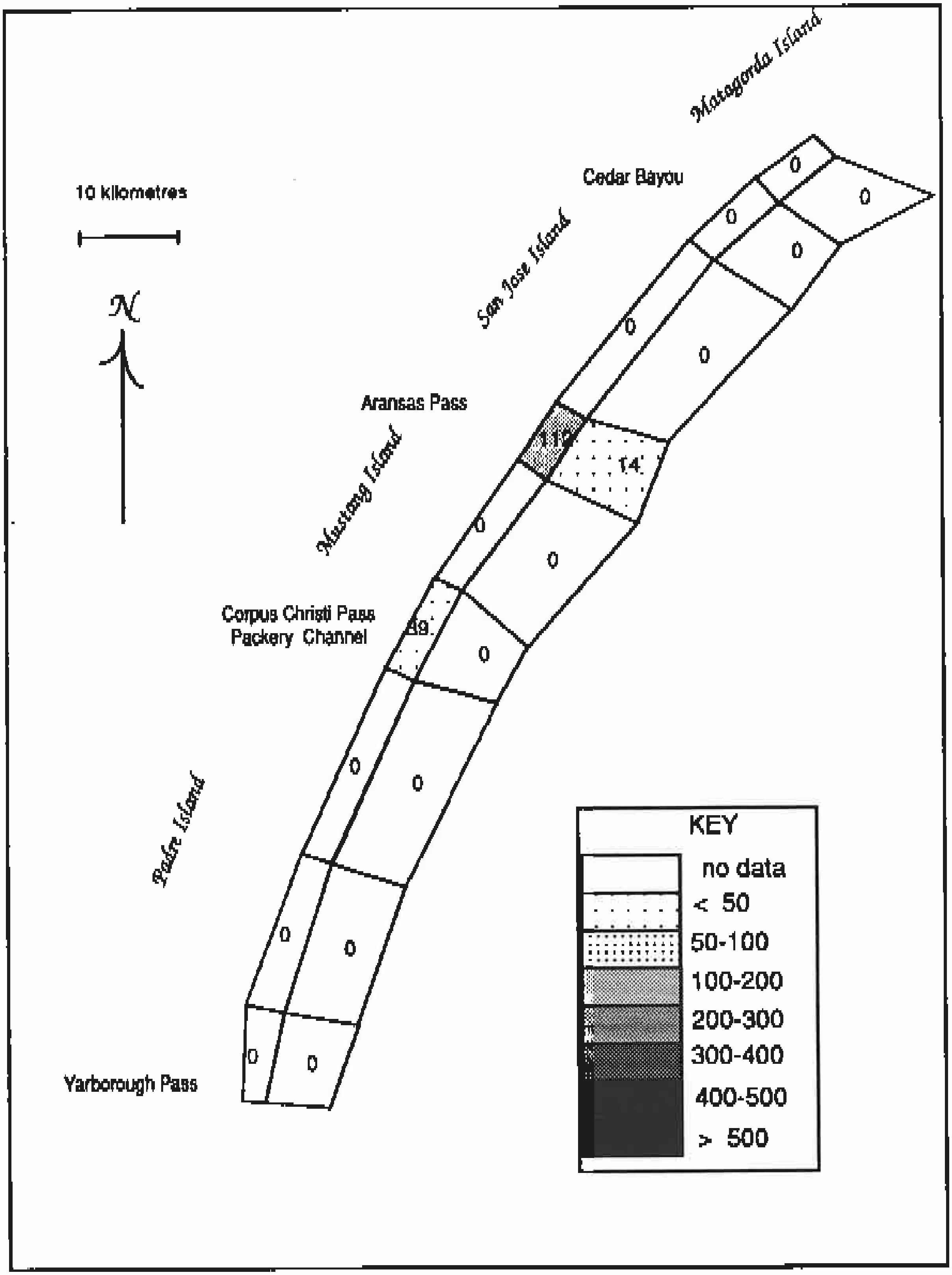


Figure 5-30. Sampling density of WQAMMN for Gulf of Mexico

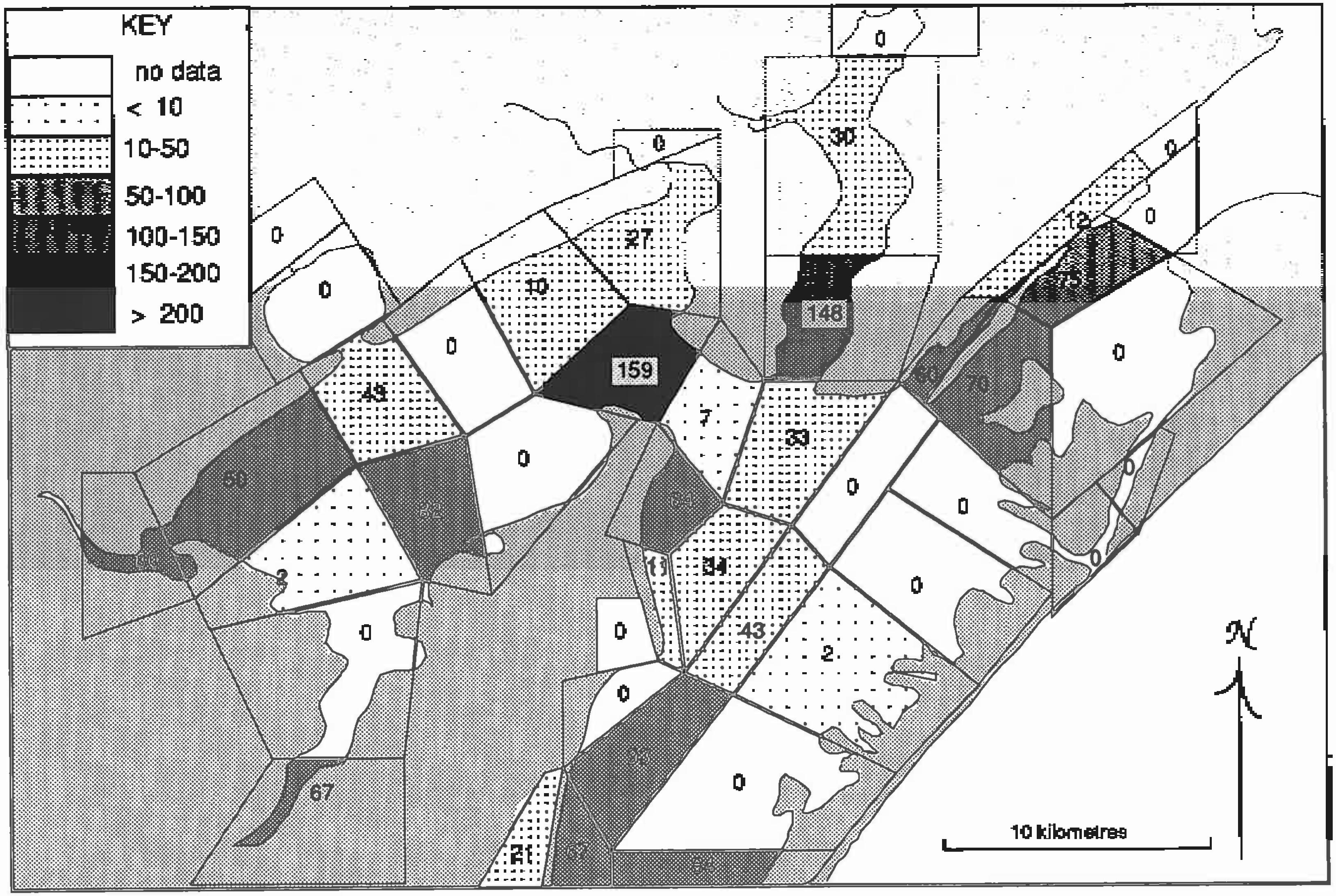


Figure 5-31. Sampling density of WQTOTP for Aransas-Copano system

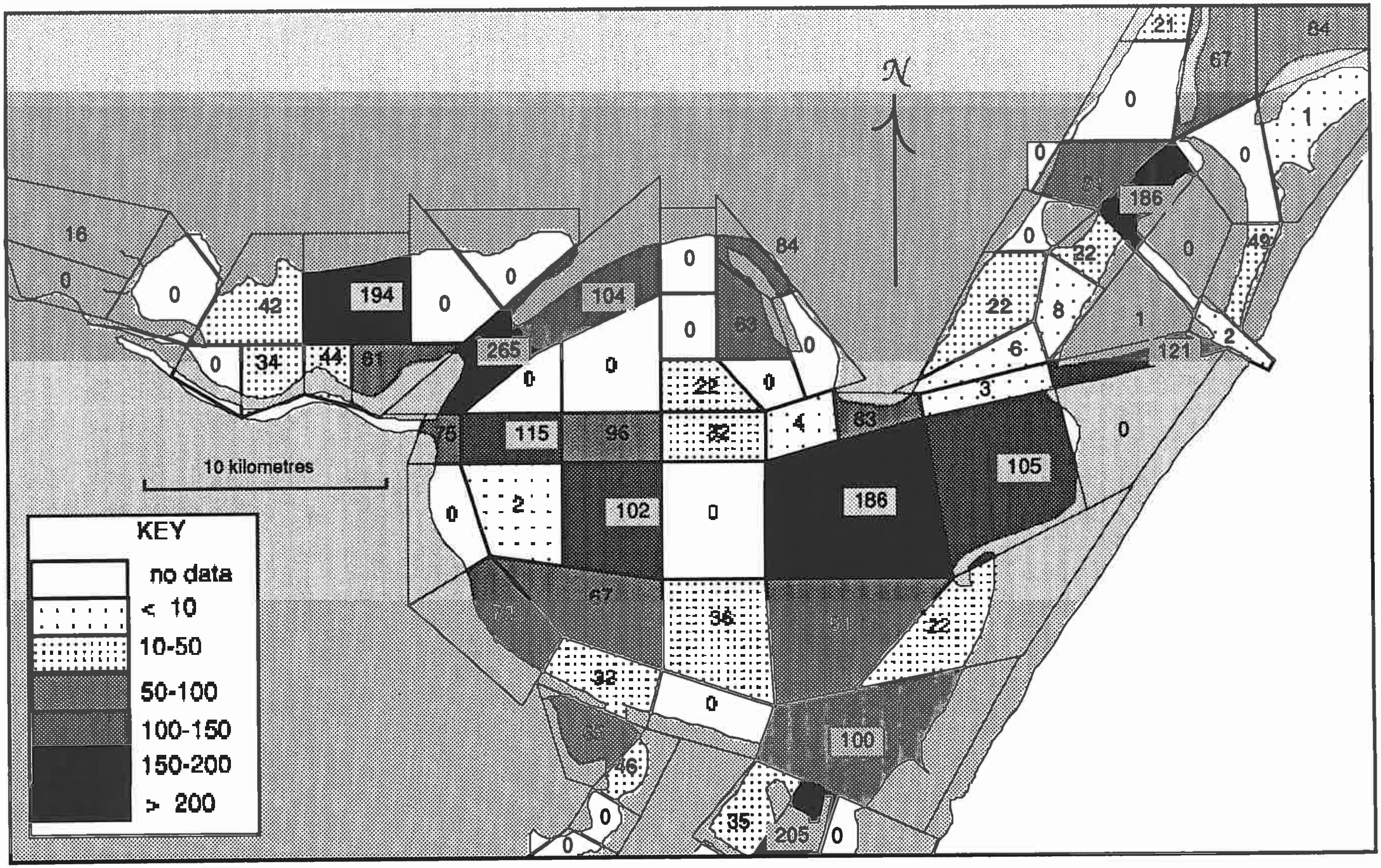


Figure 5-32. Sampling density of WQTOTP for Corpus Christi system

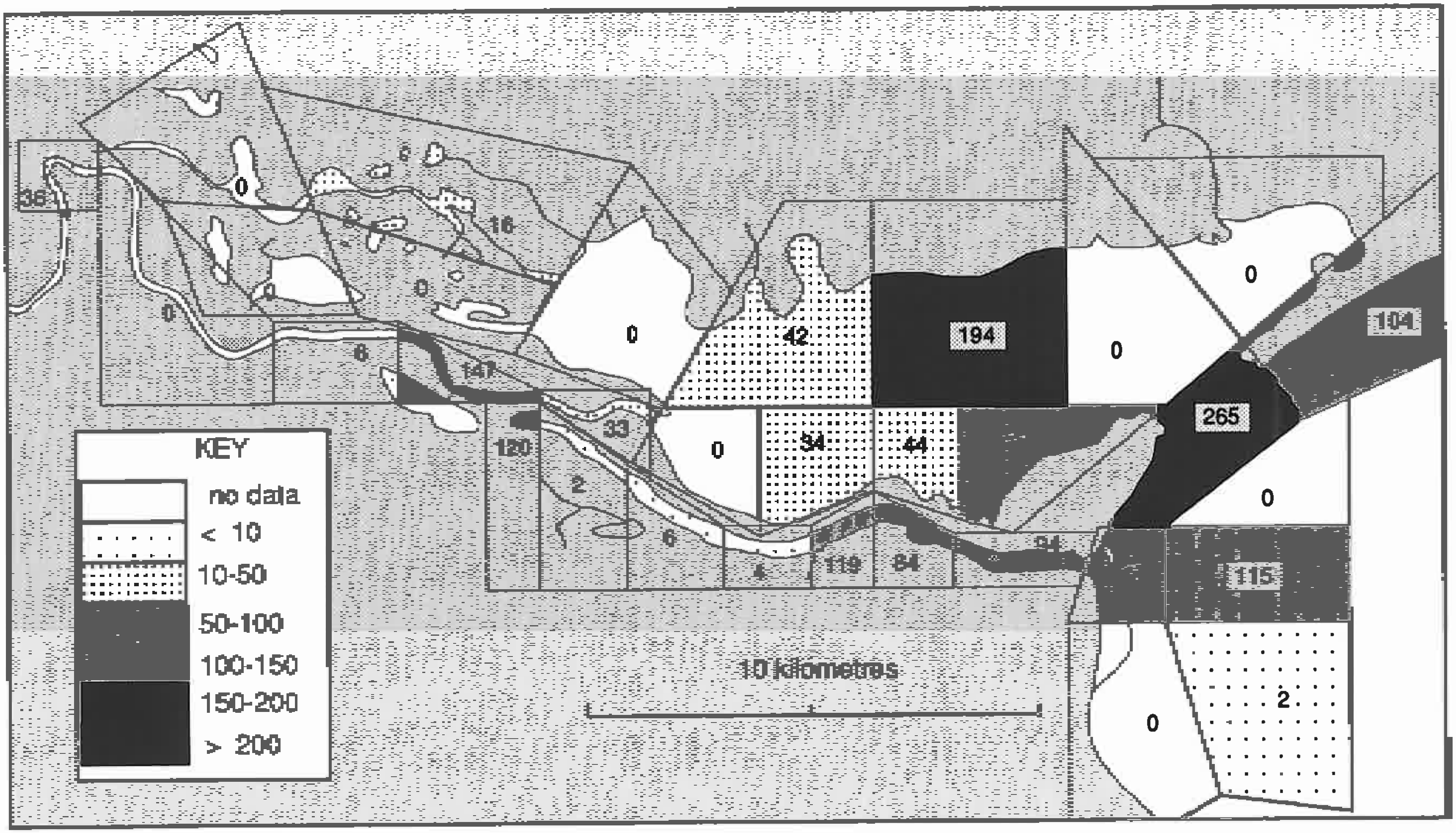


Figure 5-33. Sampling density of WQTOTP for Nueces Bay region, including Inner Harbor

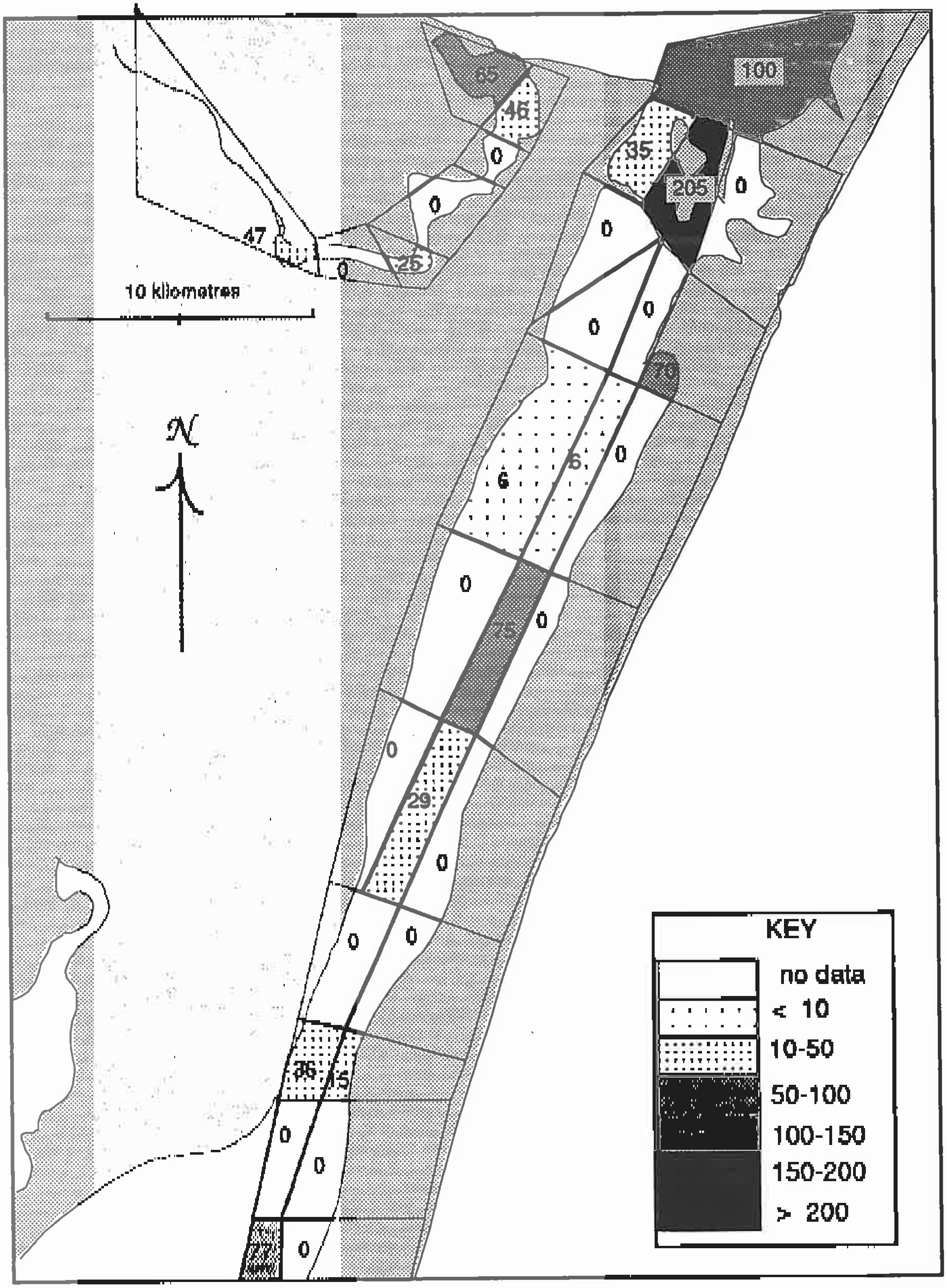


Figure 5-34. Sampling density of WQTOTP for Upper Laguna Madre and Oso Bay

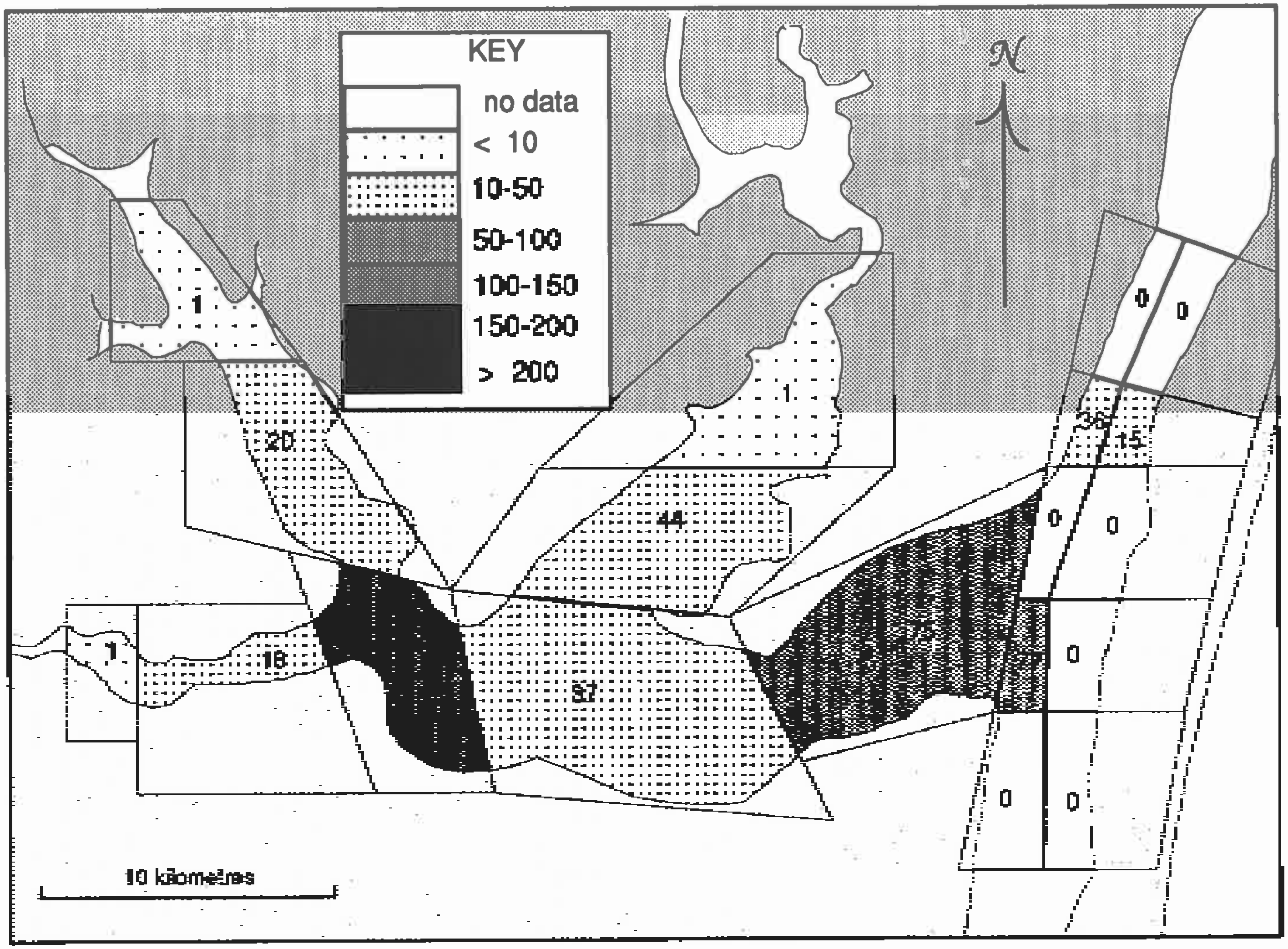


Figure 5-35. Sampling density of WQTOTP for Baffin Bay region

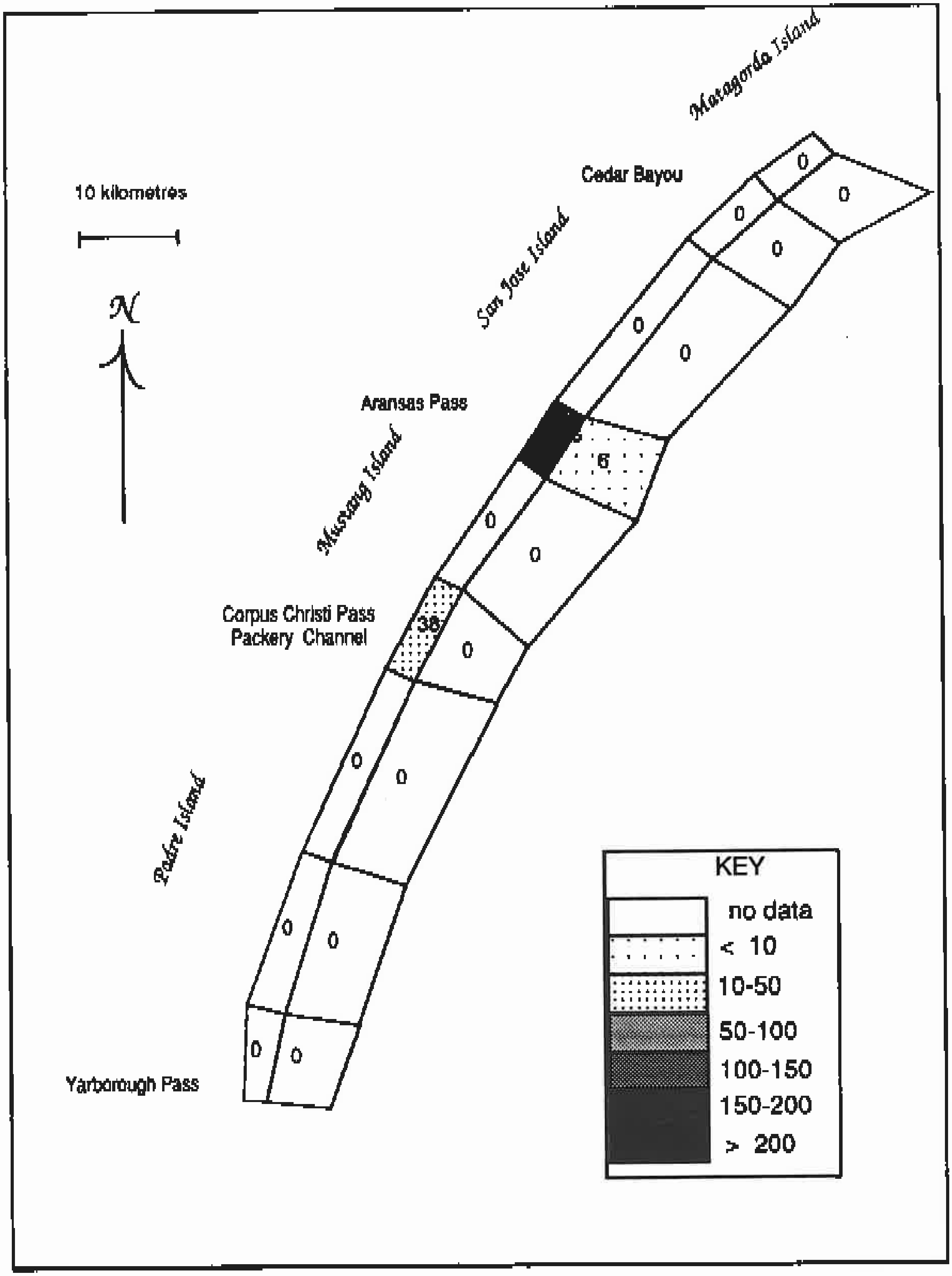


Figure 5-36. Sampling density of WQTOTP for Gulf of Mexico

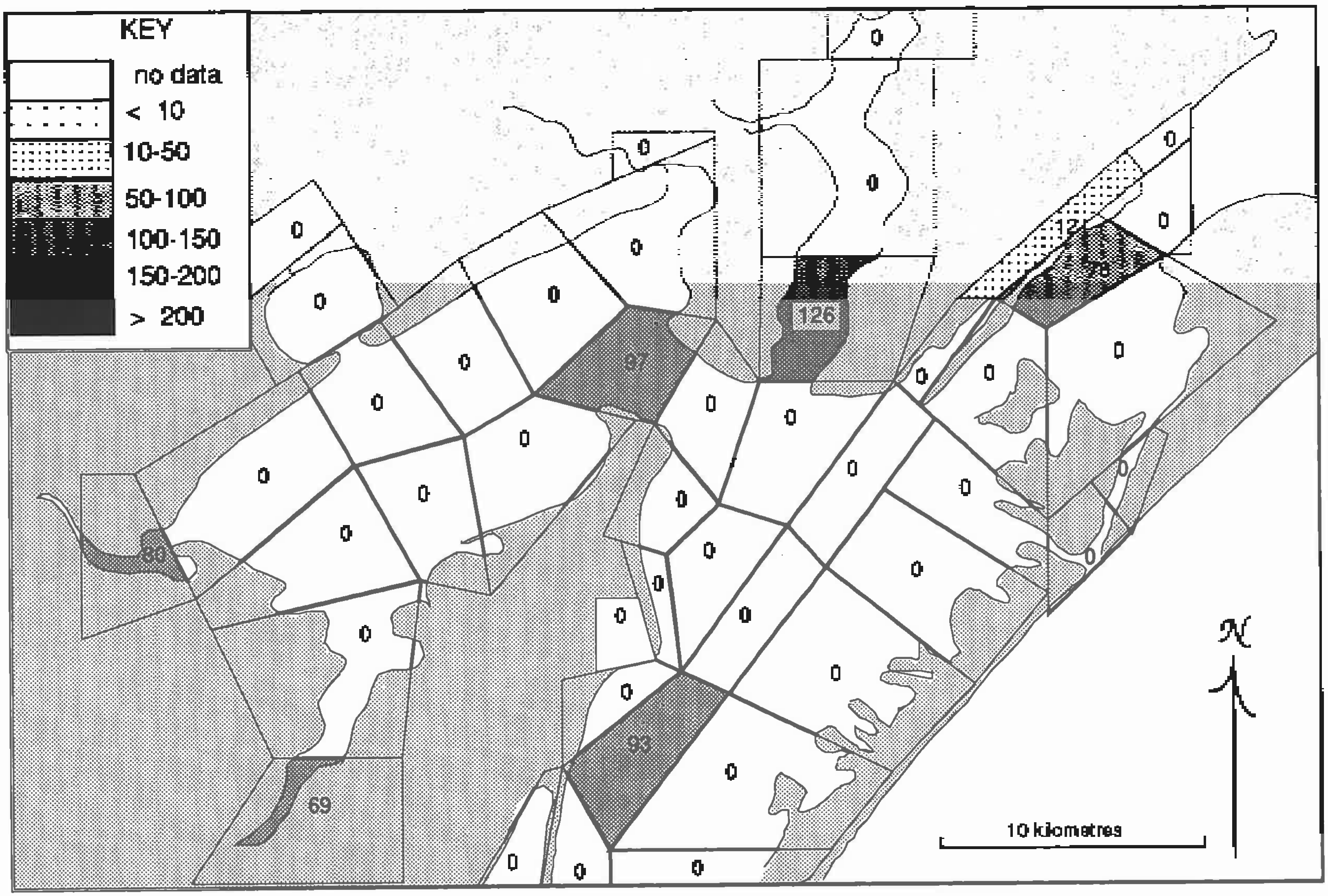


Figure 5-37. Sampling density of WQTSS for Aransas-Copano system

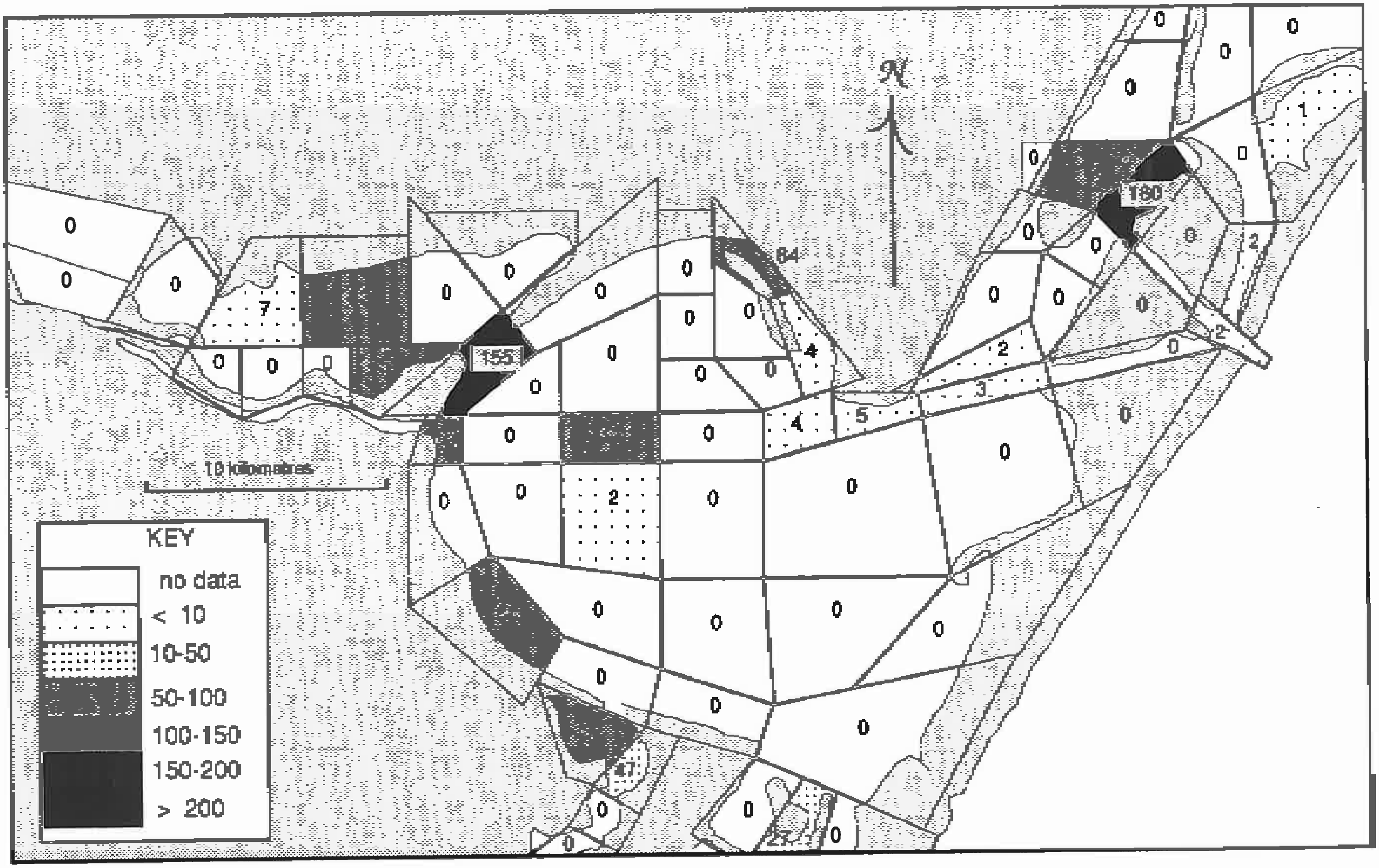


Figure 5-38. Sampling density of WQTSS for Corpus Christi system

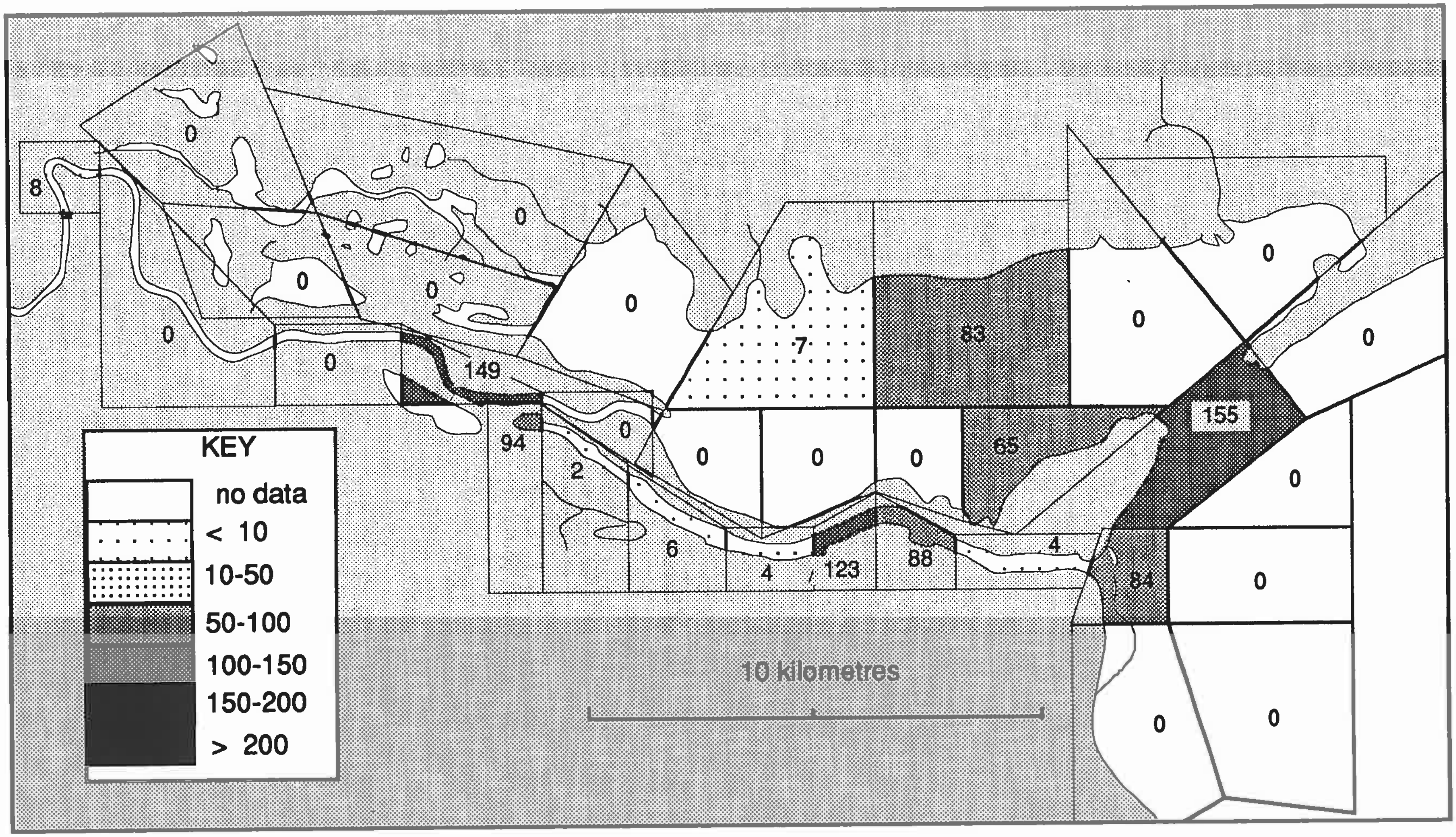


Figure 5-39. Sampling density of WQTSS for Nueces Bay region, including Inner Harbor

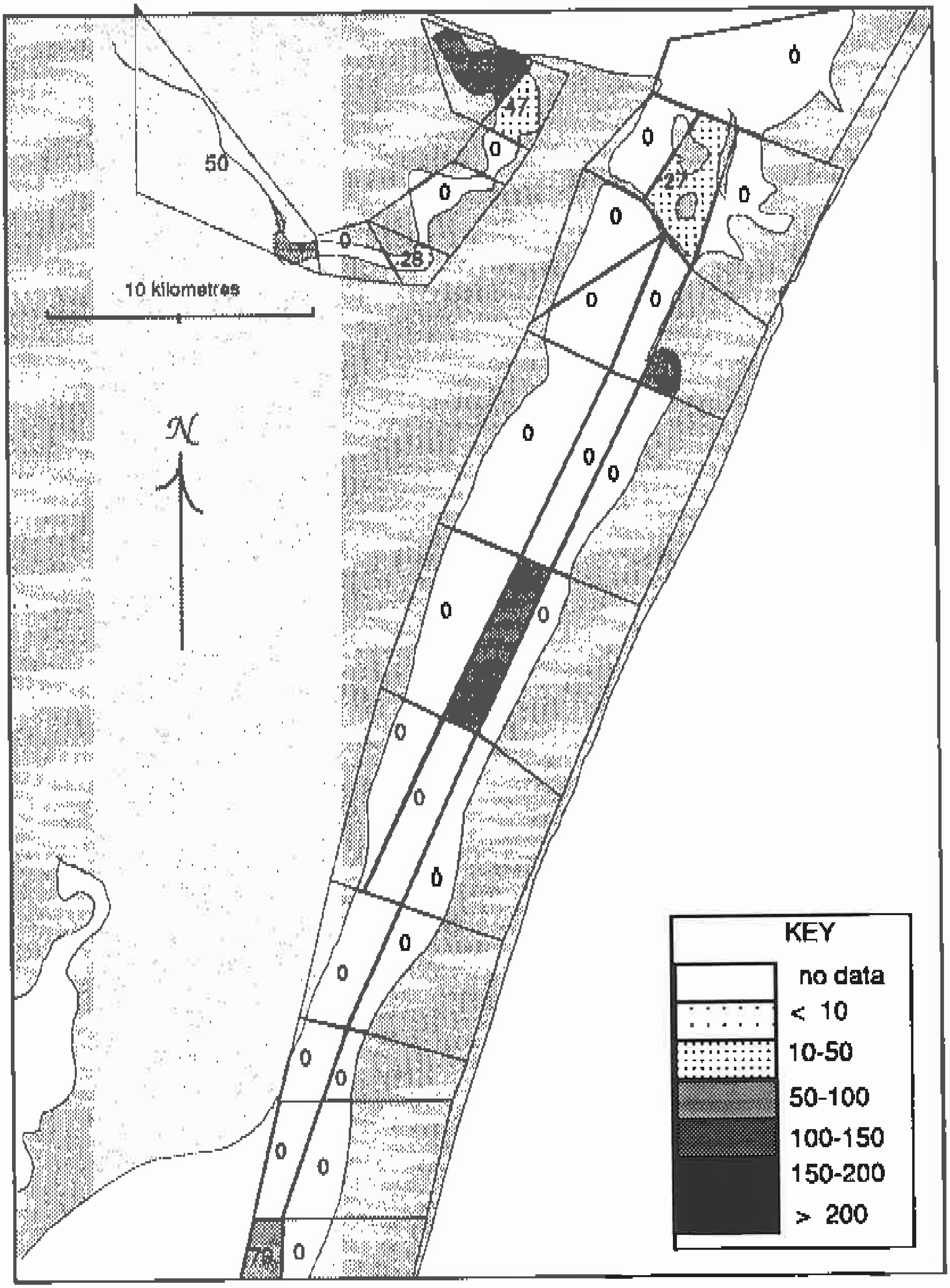


Figure 5-40. Sampling density of WQTSS for Upper Laguna Madre and Oso Bay

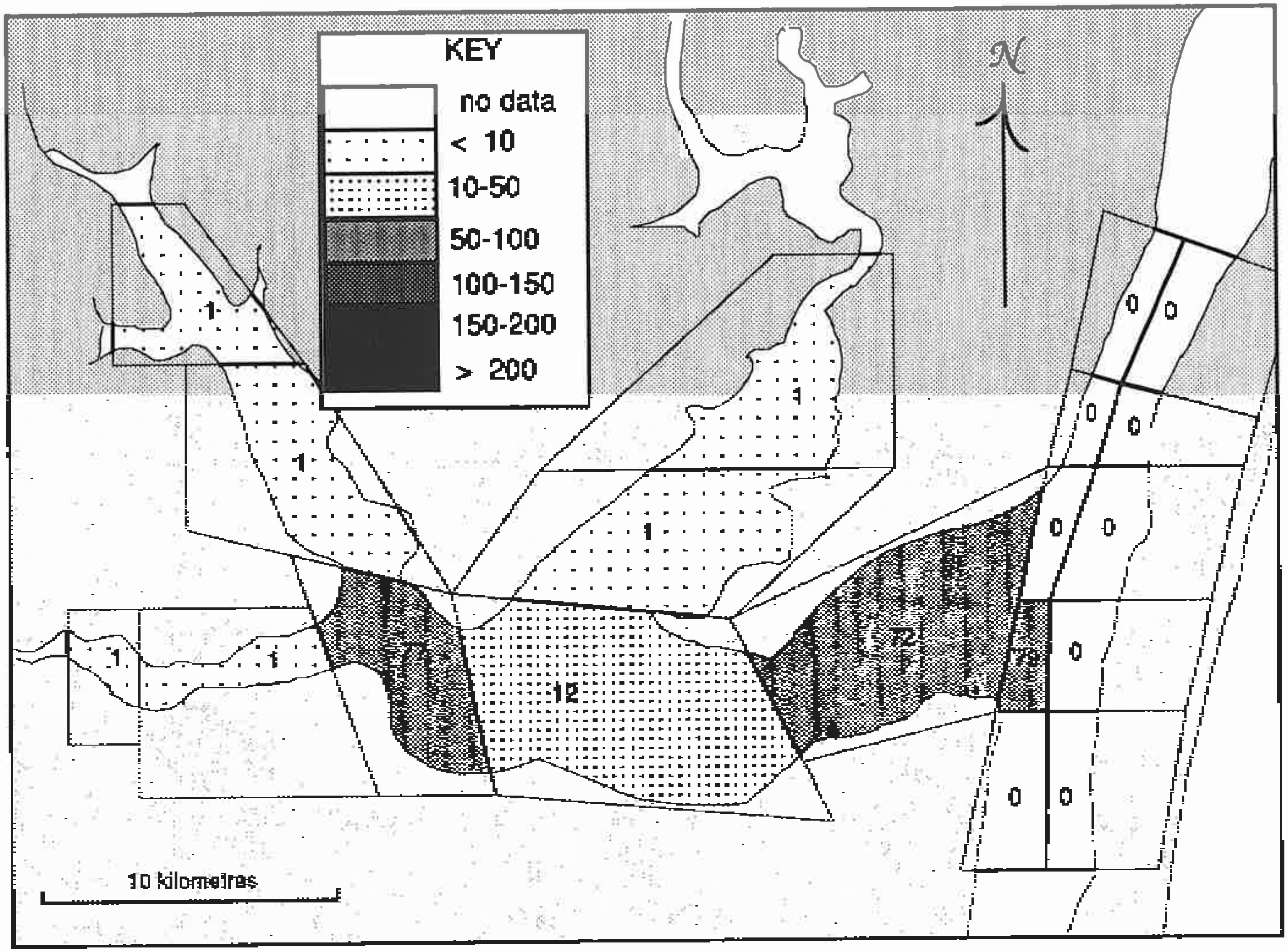


Figure 5-41. Sampling density of WQTSS for Baffin Bay region

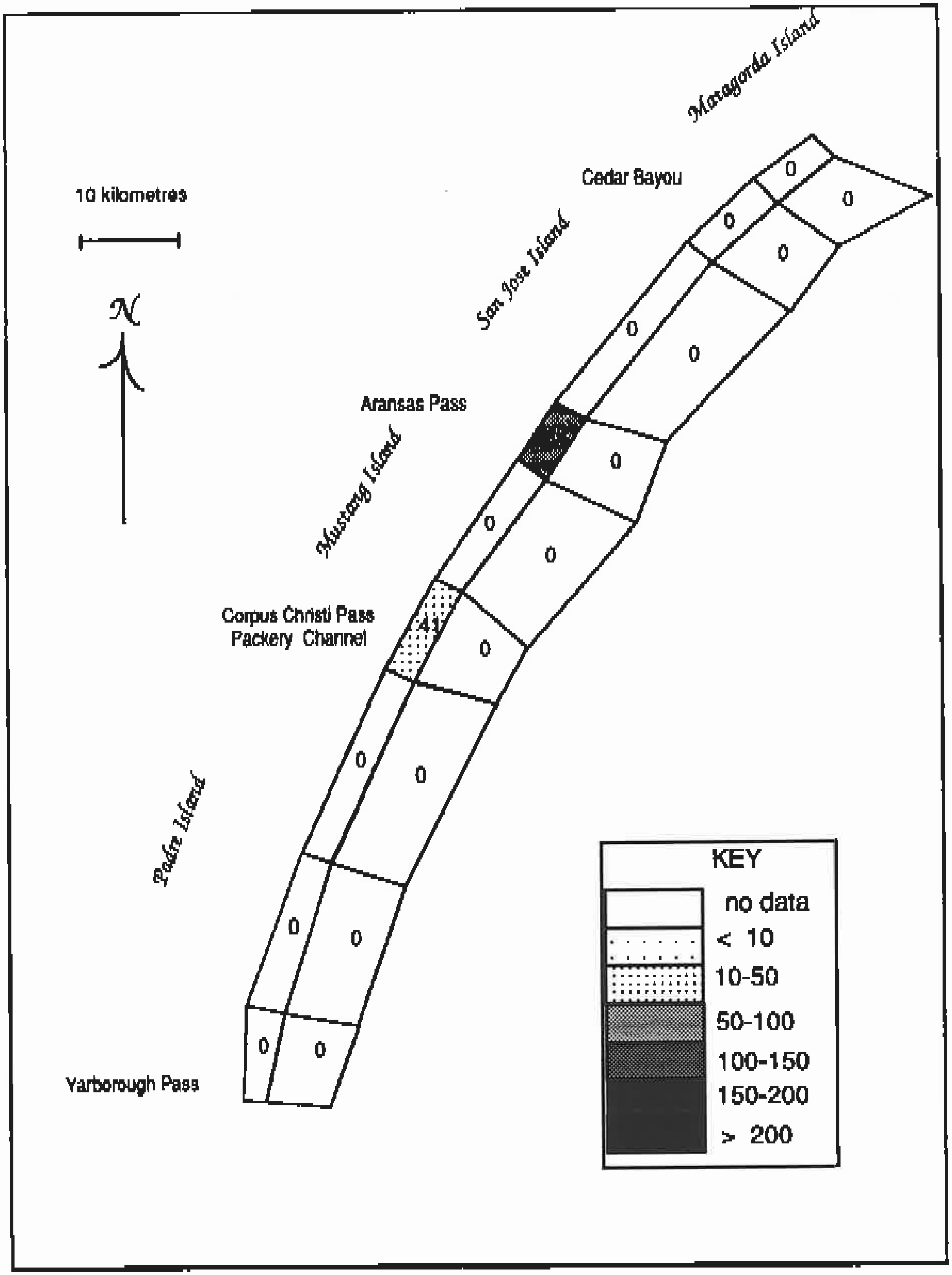


Figure 5-42. Sampling density of WQTSS for Gulf of Mexico

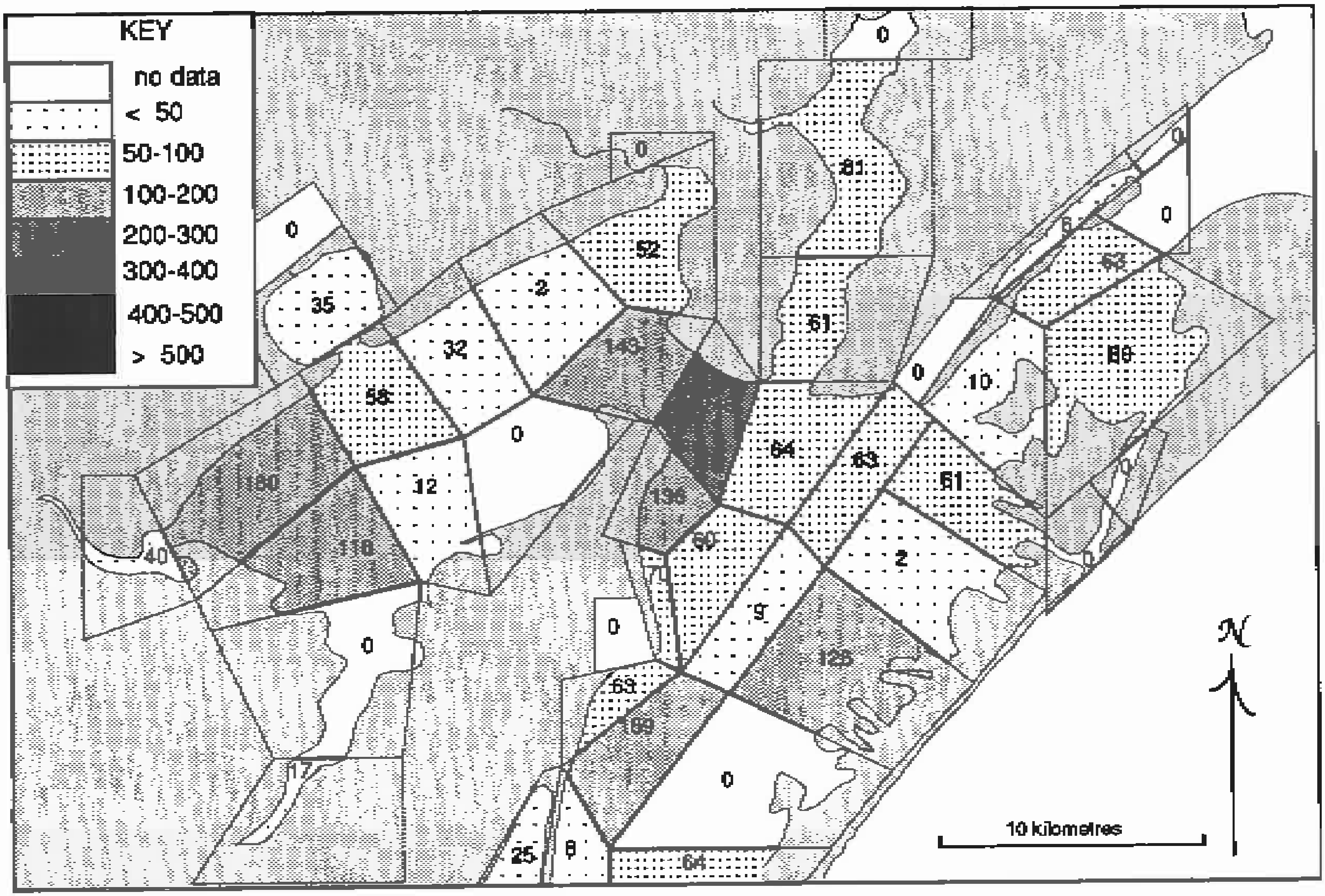


Figure 5-43. Sampling density of WQFCOLI for Aransas-Copano system

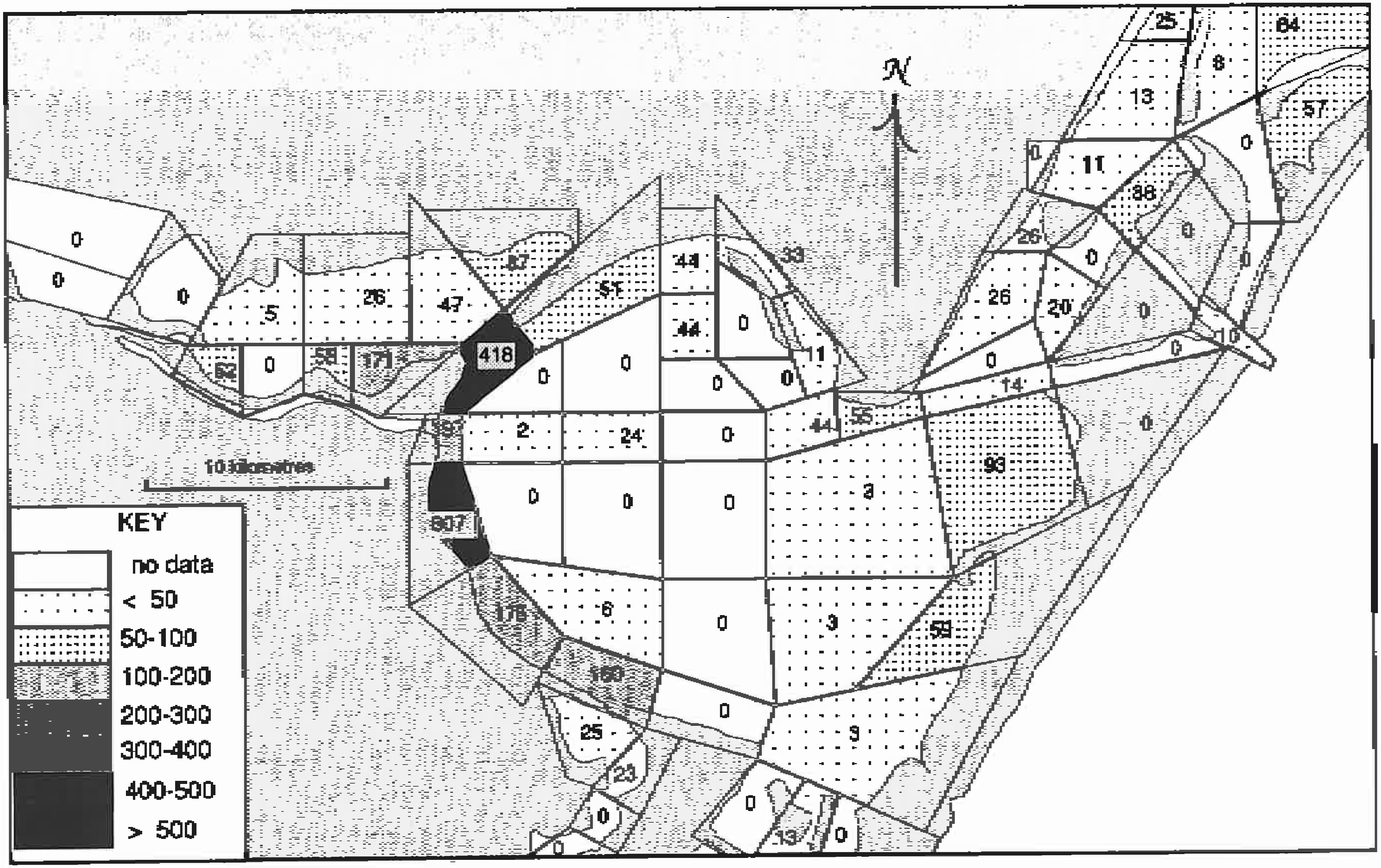


Figure 5-44. Sampling density of WQFCOLI for Corpus Christi system

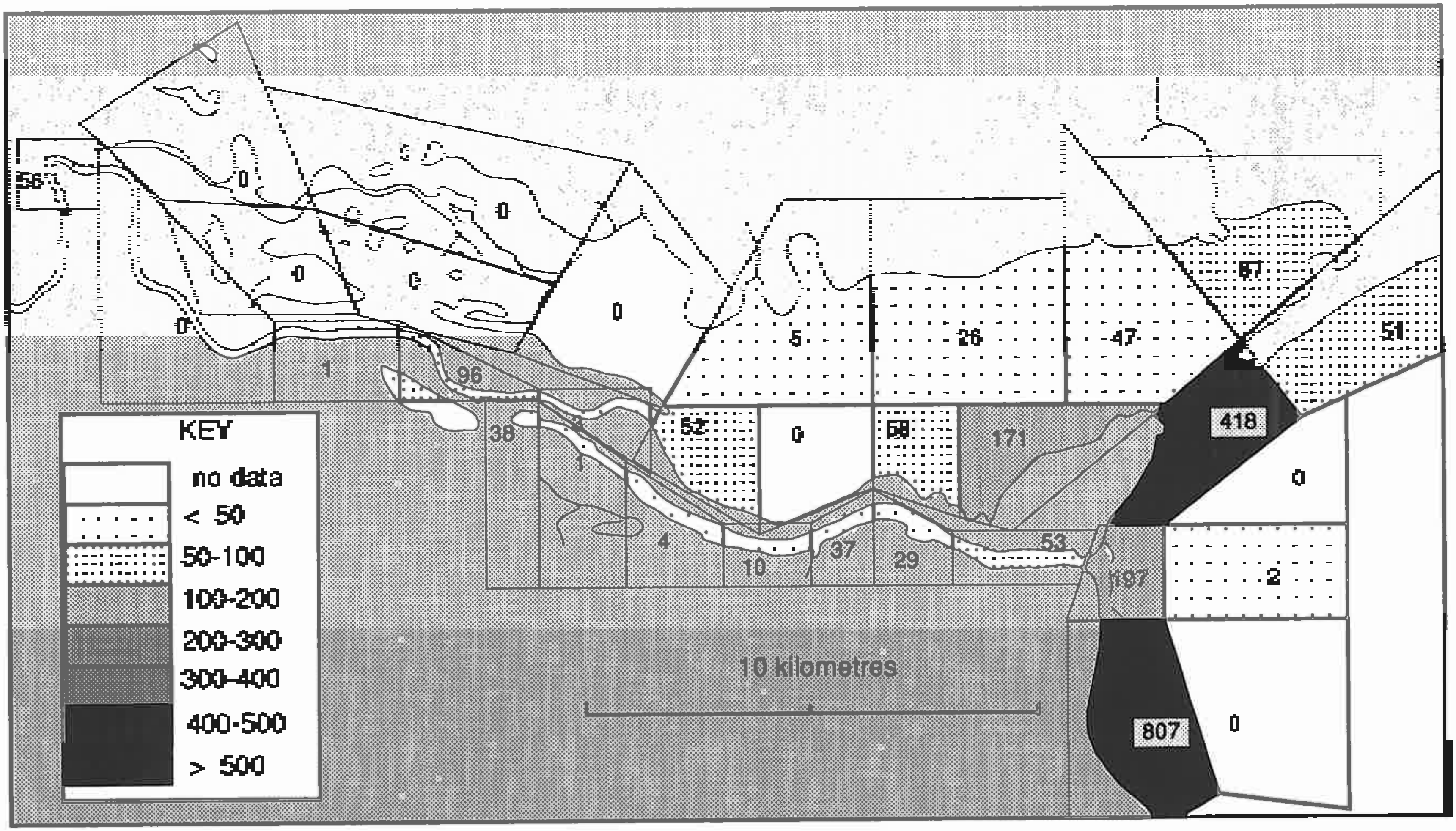


Figure 5-45. Sampling density of WQFCOLI for Nueces Bay region, including Inner Harbor

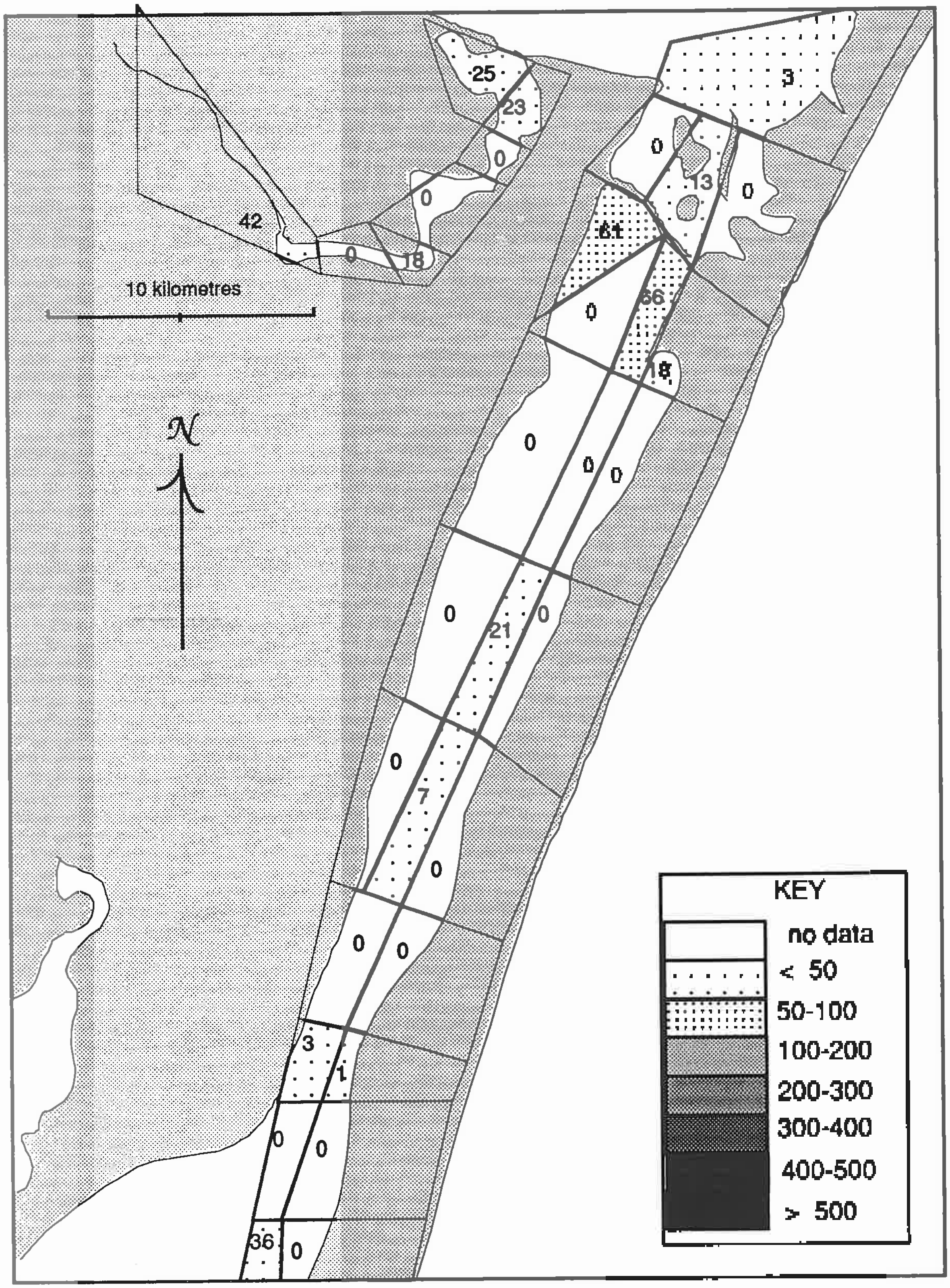


Figure 5-46. Sampling density of WQFCOLI for Upper Laguna Madre and Oso Bay

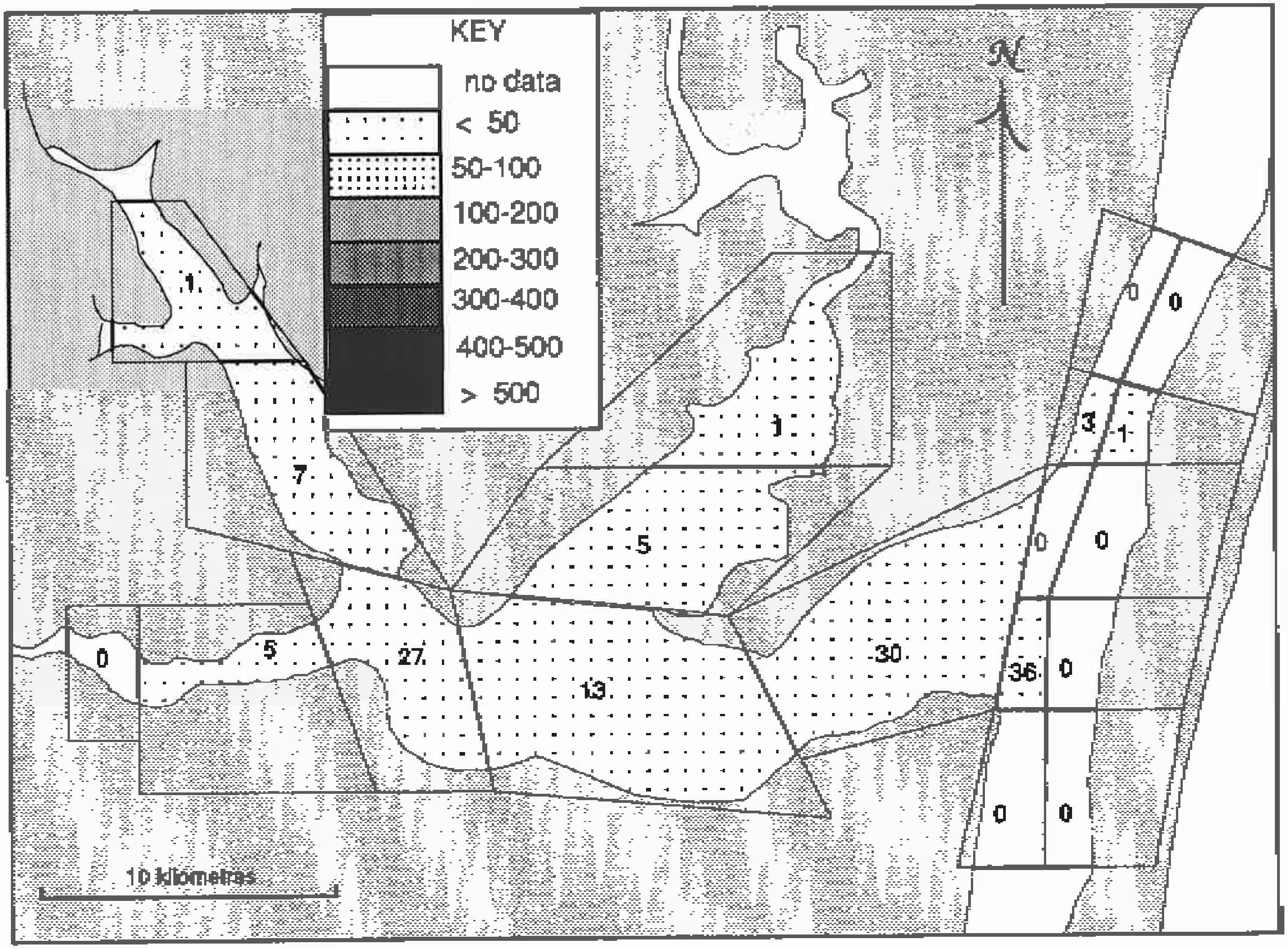


Figure 5-47. Sampling density of WQFCOLI for Baffin Bay region

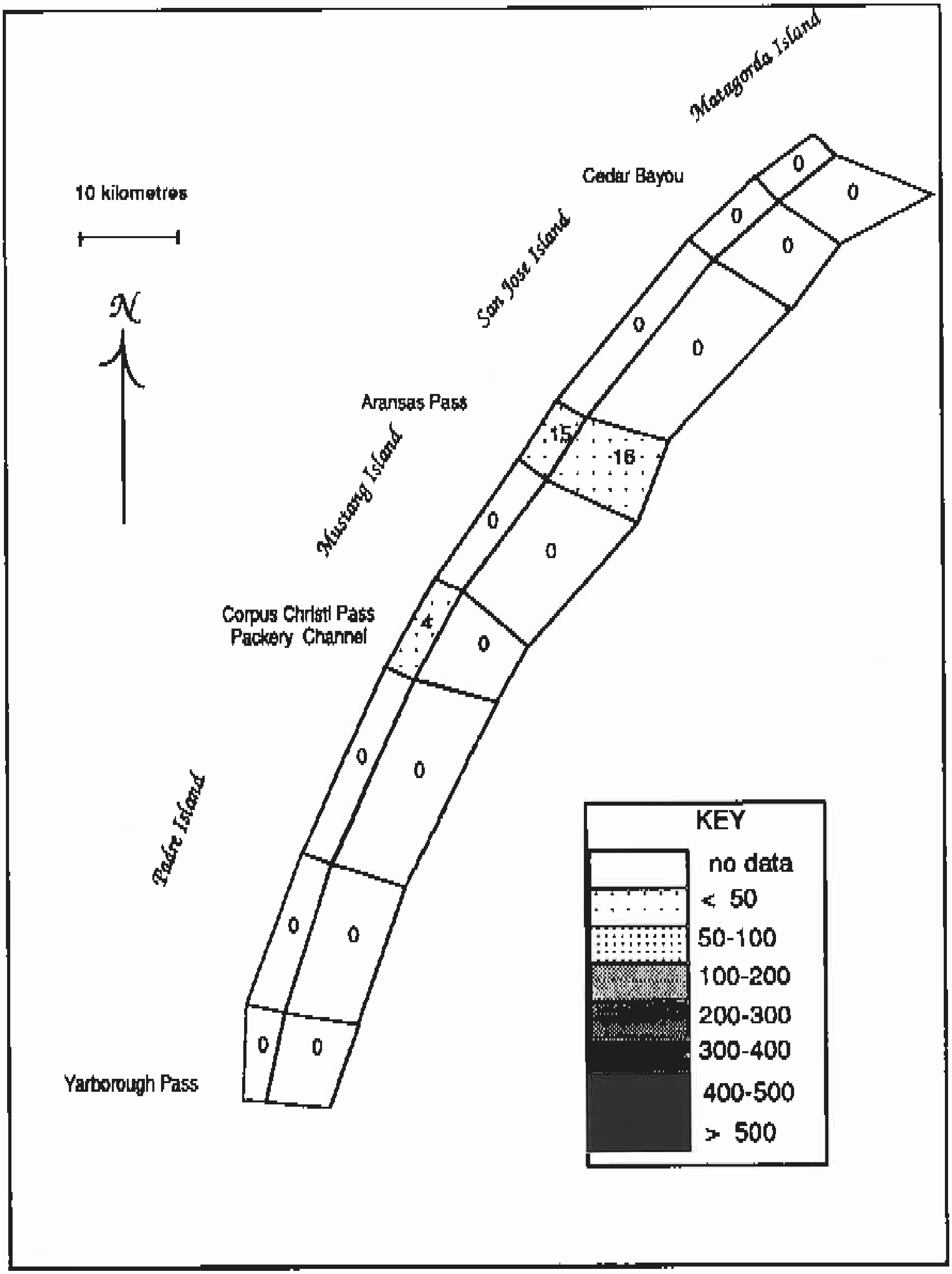


Figure 5-48. Sampling density of WQFCOLI for Gulf of Mexico

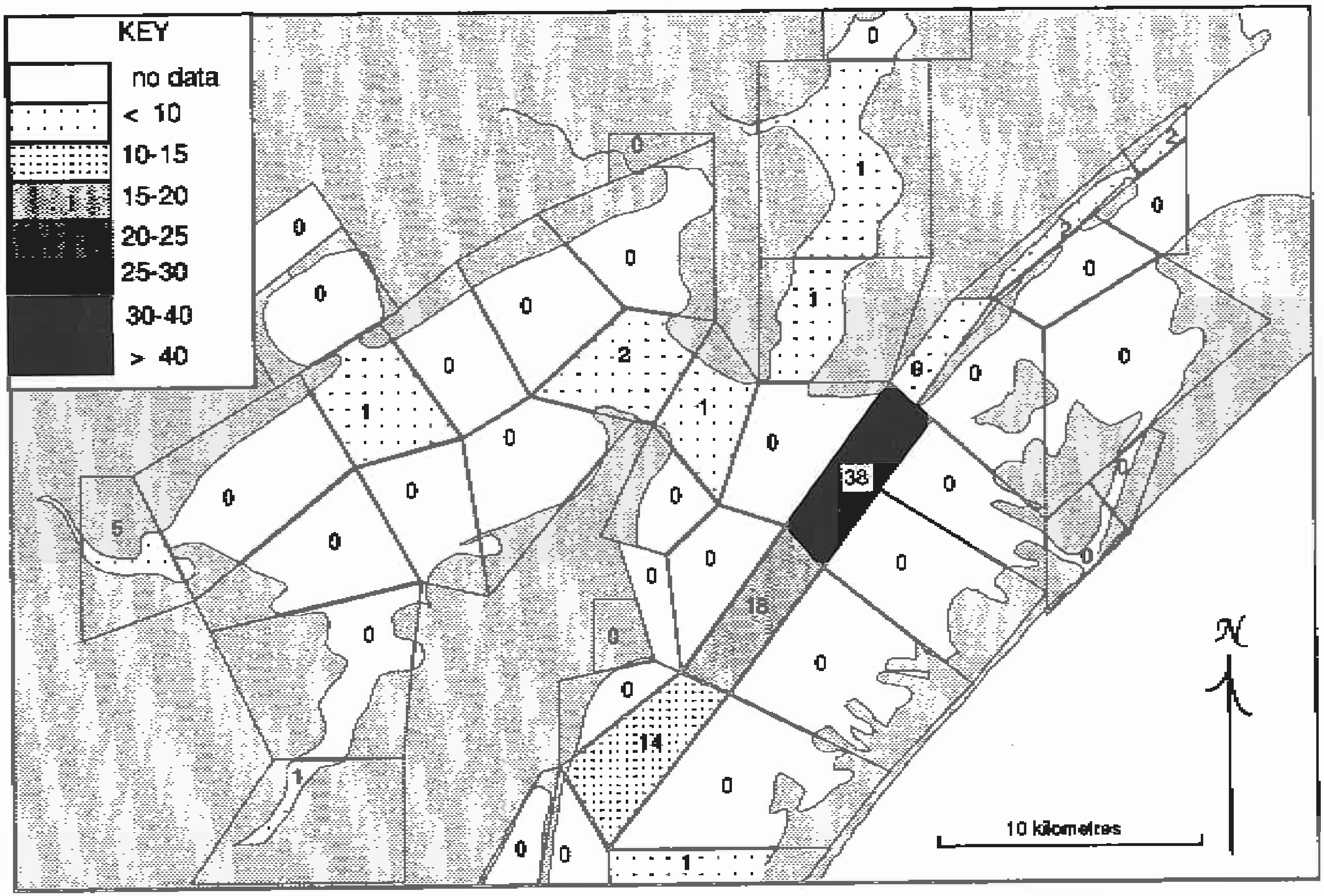


Figure 5-49. Sampling density of WQMETCDT for Aransas-Copano system

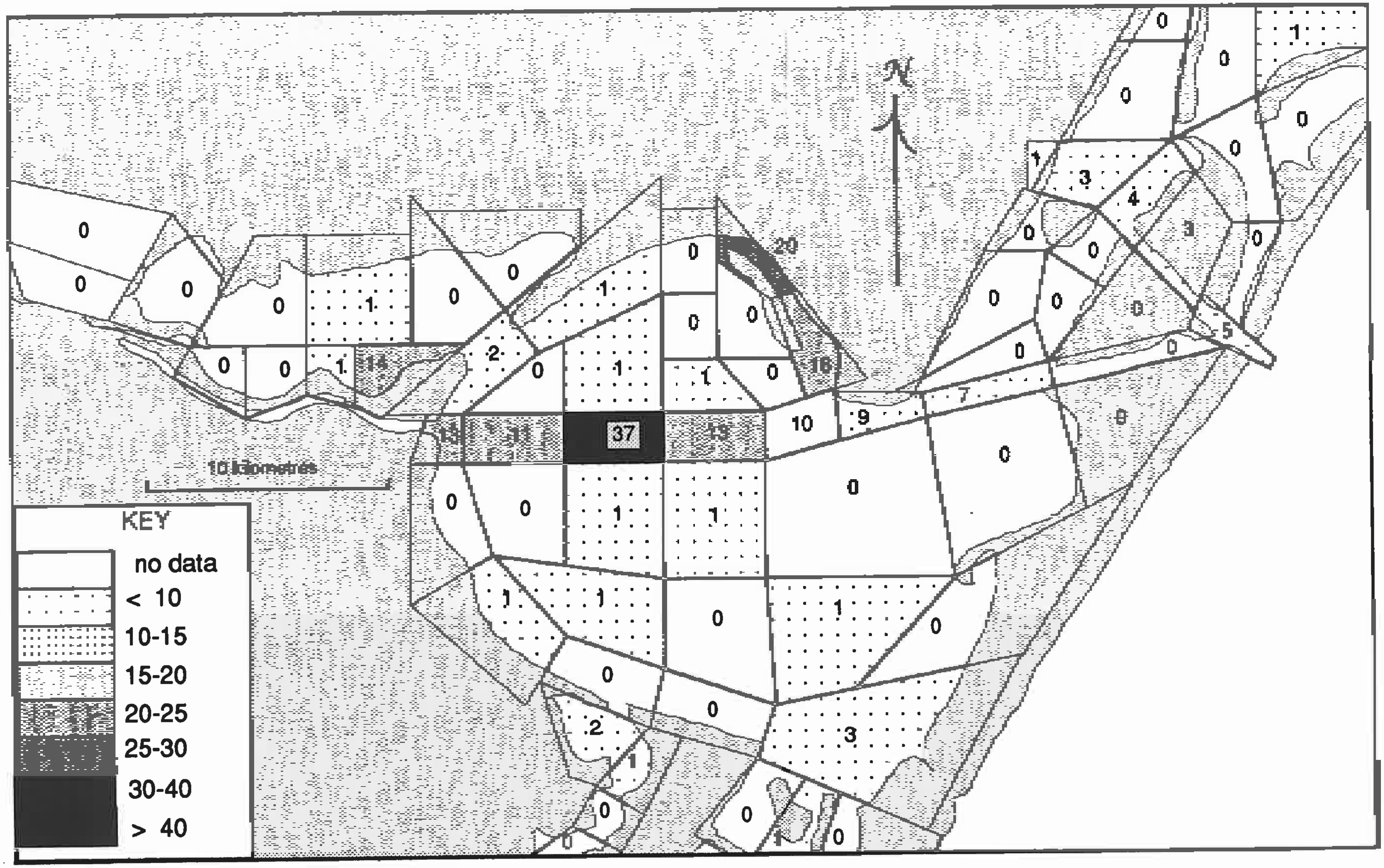


Figure 5-50. Sampling density of WQMETCDT for Corpus Christi system

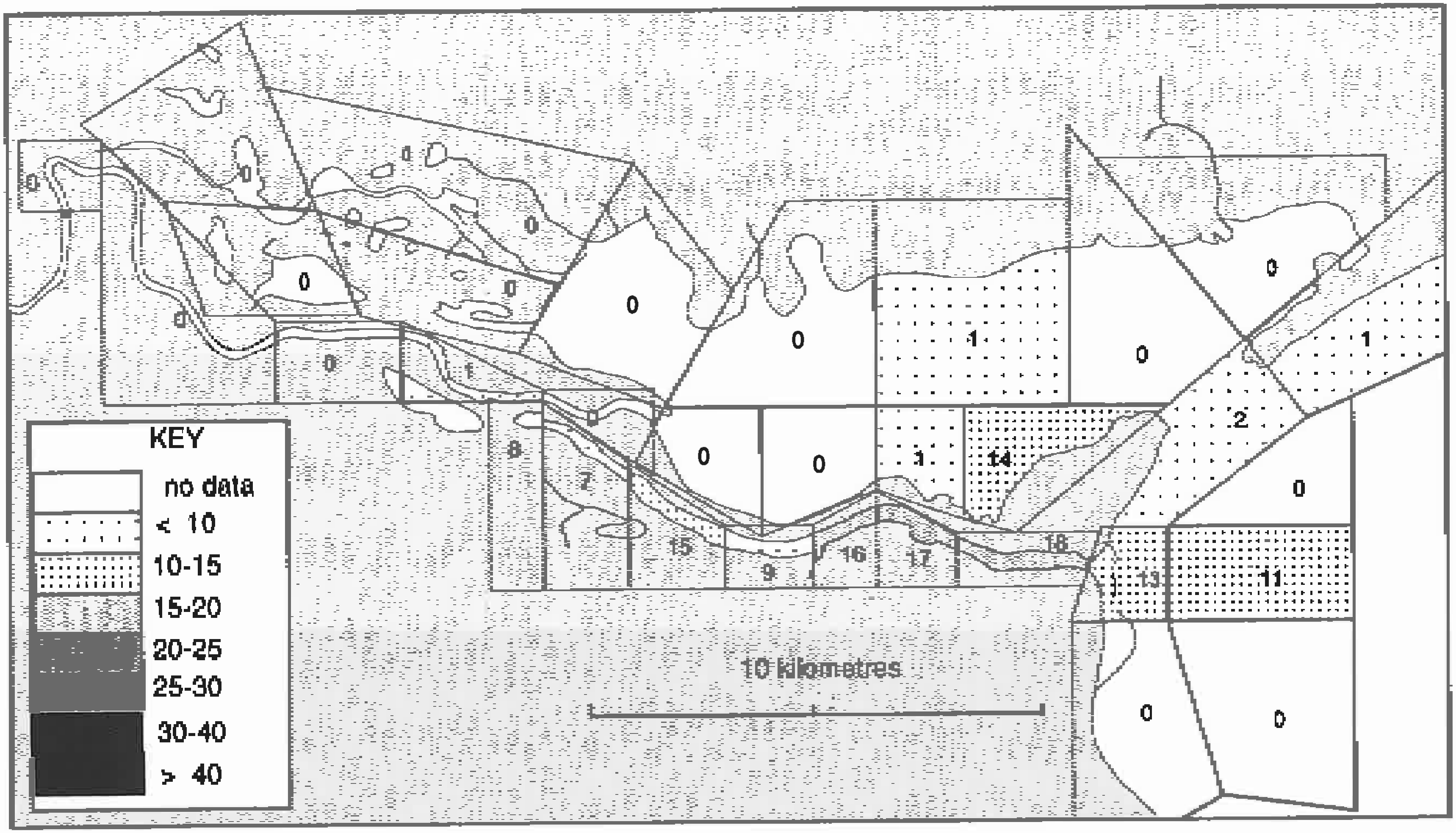


Figure 5-51. Sampling density of WQMETCDT for Nueces Bay region, including Inner Harbor

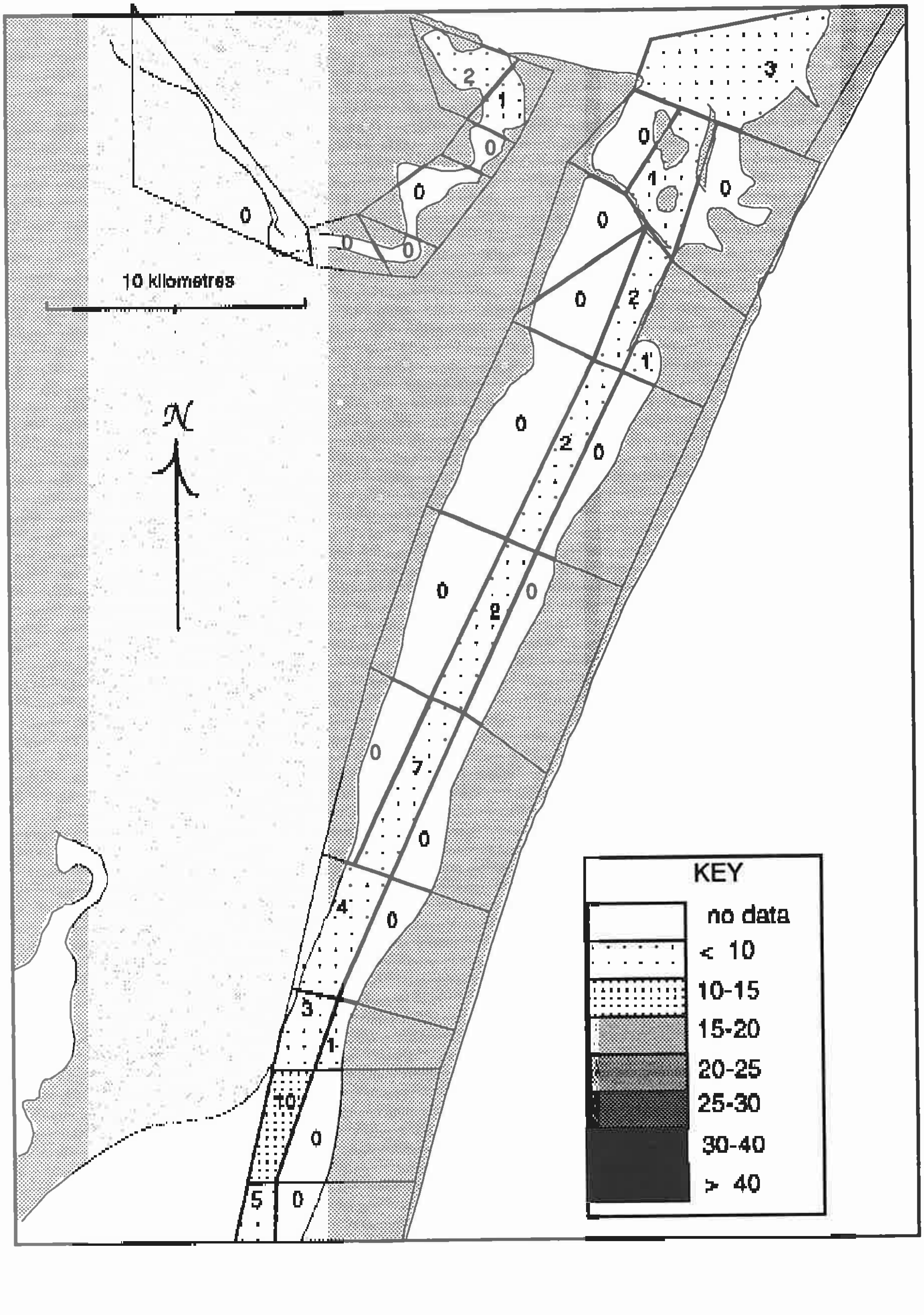


Figure 5-52. Sampling density of WQMETCDT for Upper Laguna Madre and Oso Bay

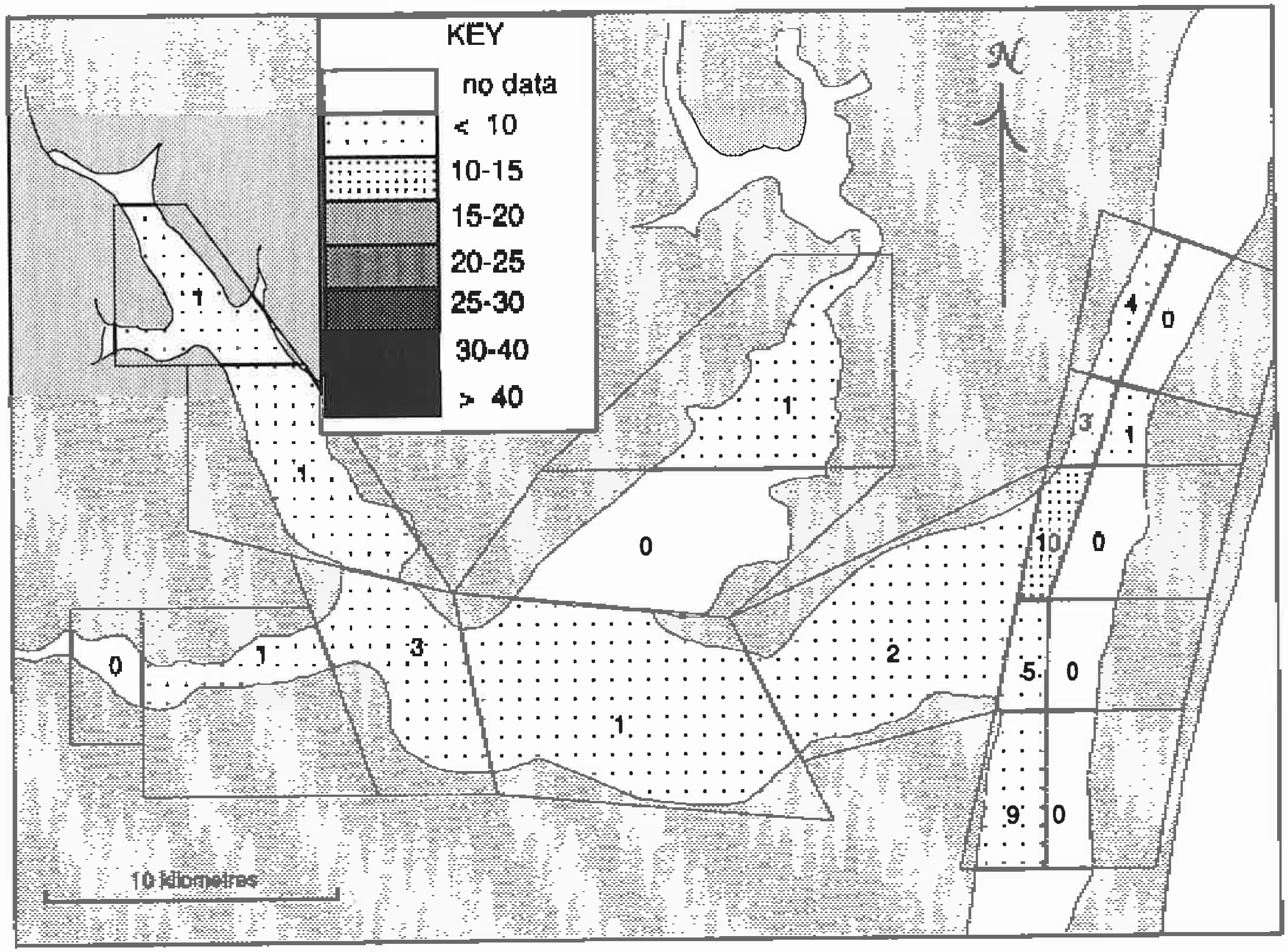


Figure 5-53. Sampling density of WQMETCDT for Baffin Bay region

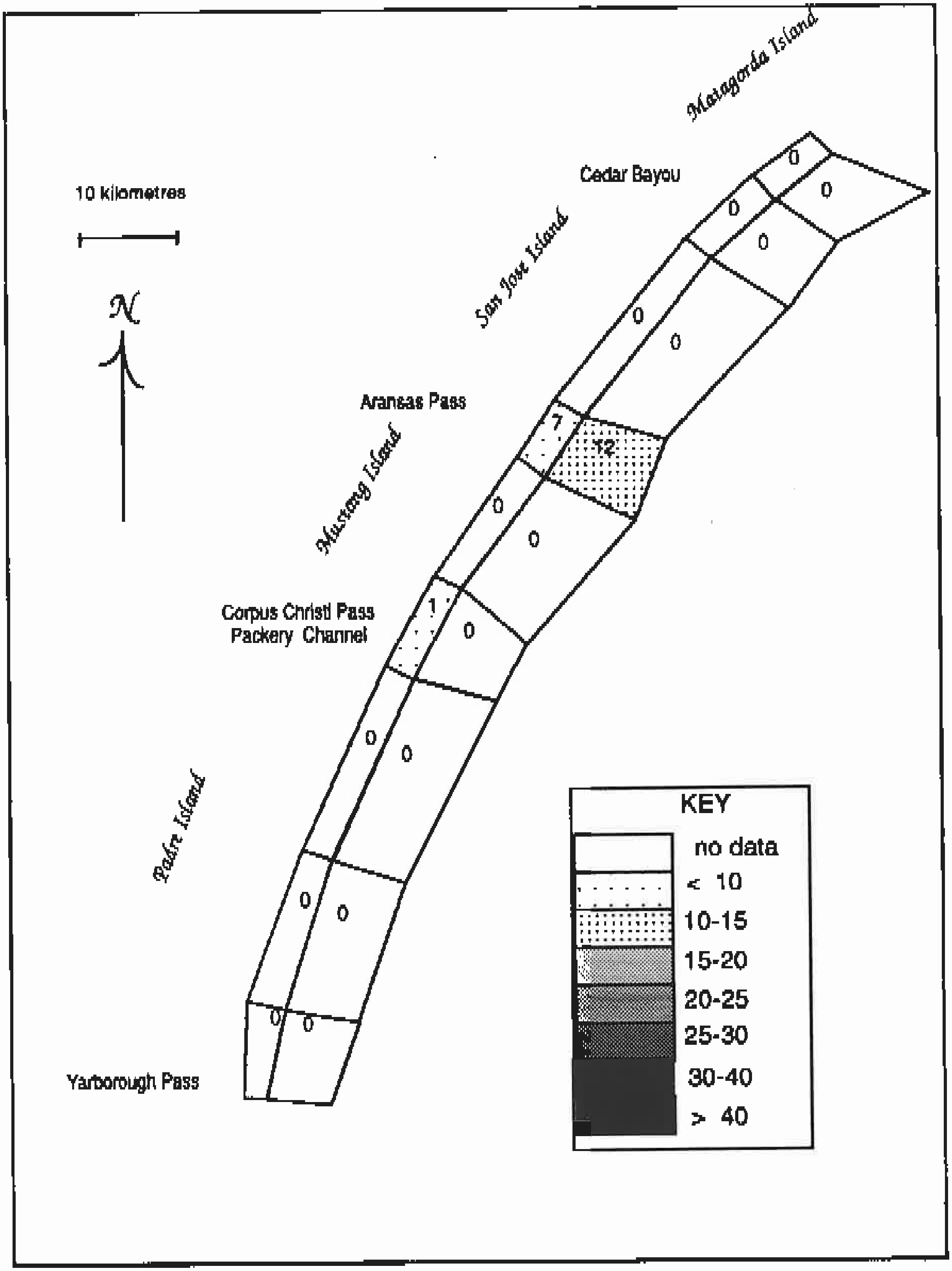


Figure 5-54. Sampling density of WQMETCDT for Gulf of Mexico

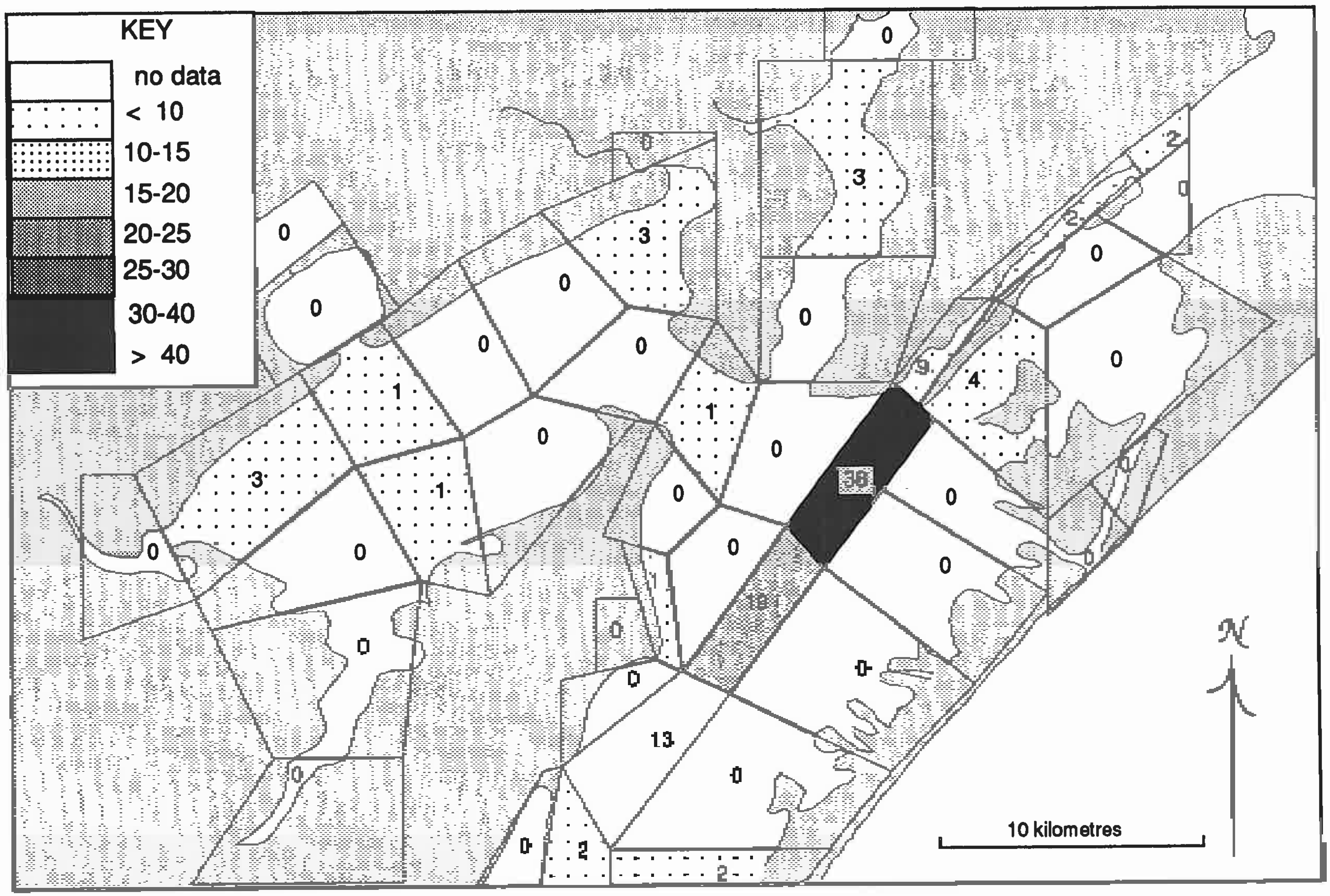


Figure 5-55. Sampling density of WQ-XDDT for Aransas-Copano system

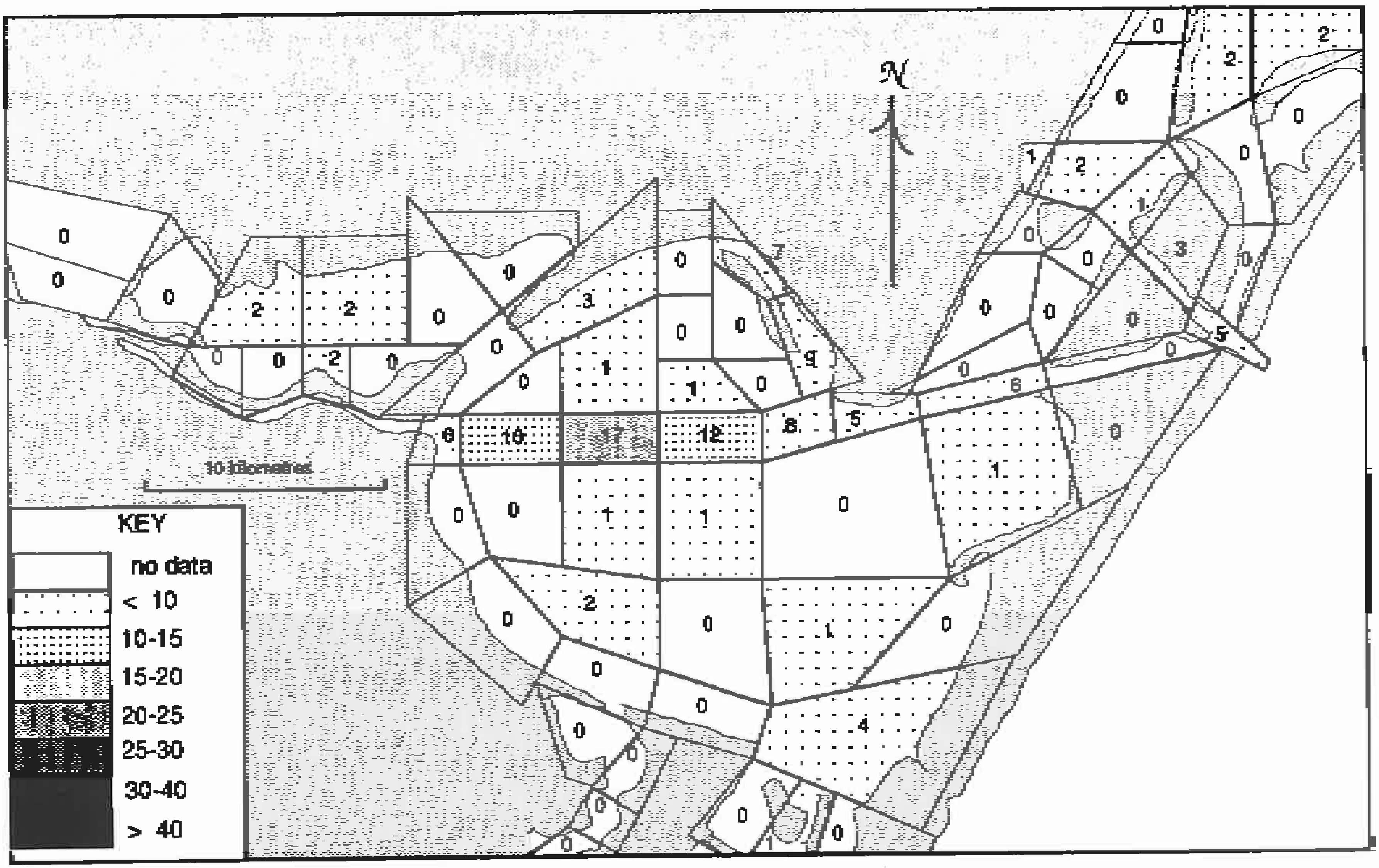


Figure 5-56. Sampling density of WQ-XDDT for Corpus Christi system

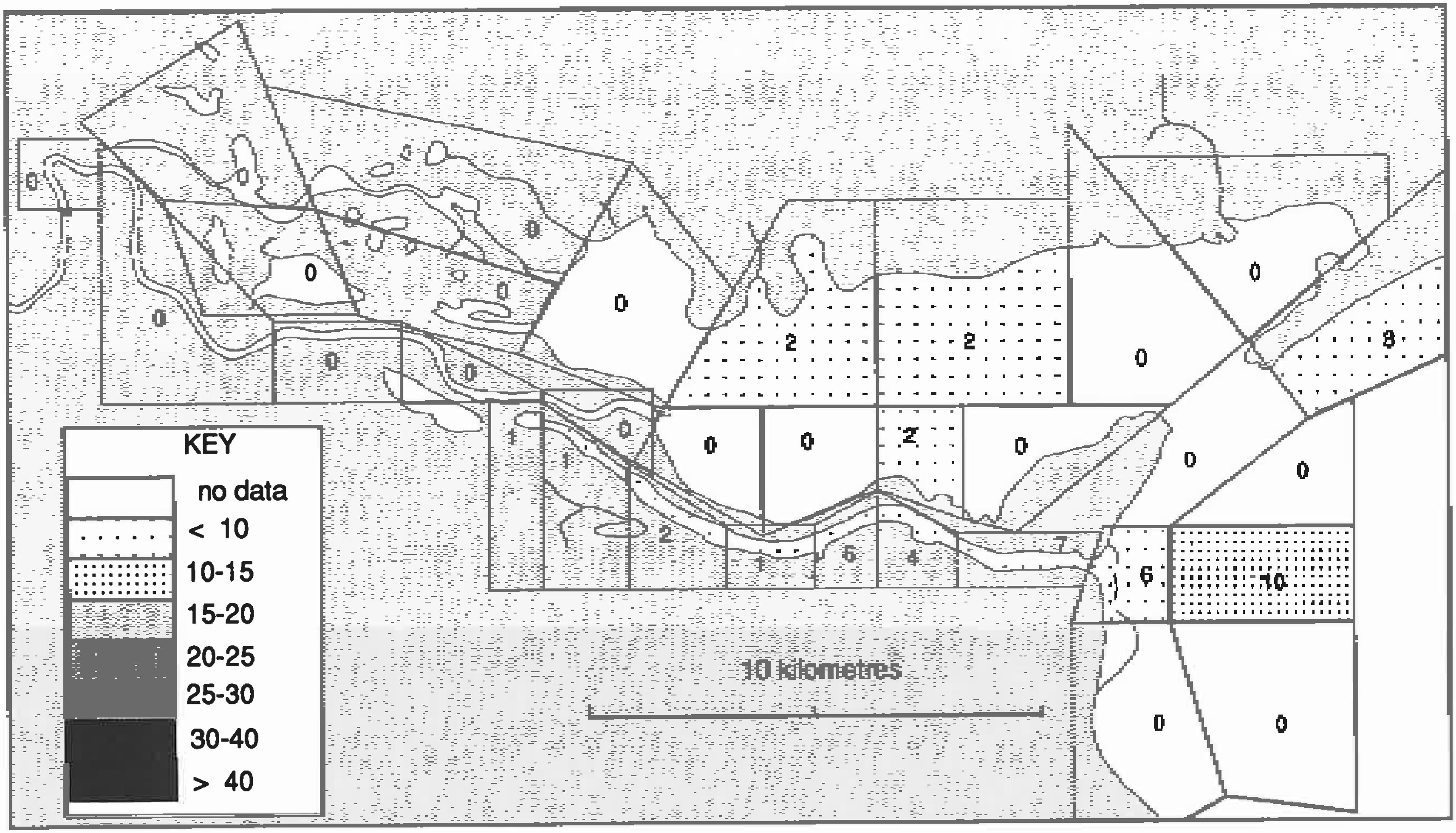


Figure 5-57. Sampling density of WQ-XDDT for Nueces Bay region, including Inner Harbor

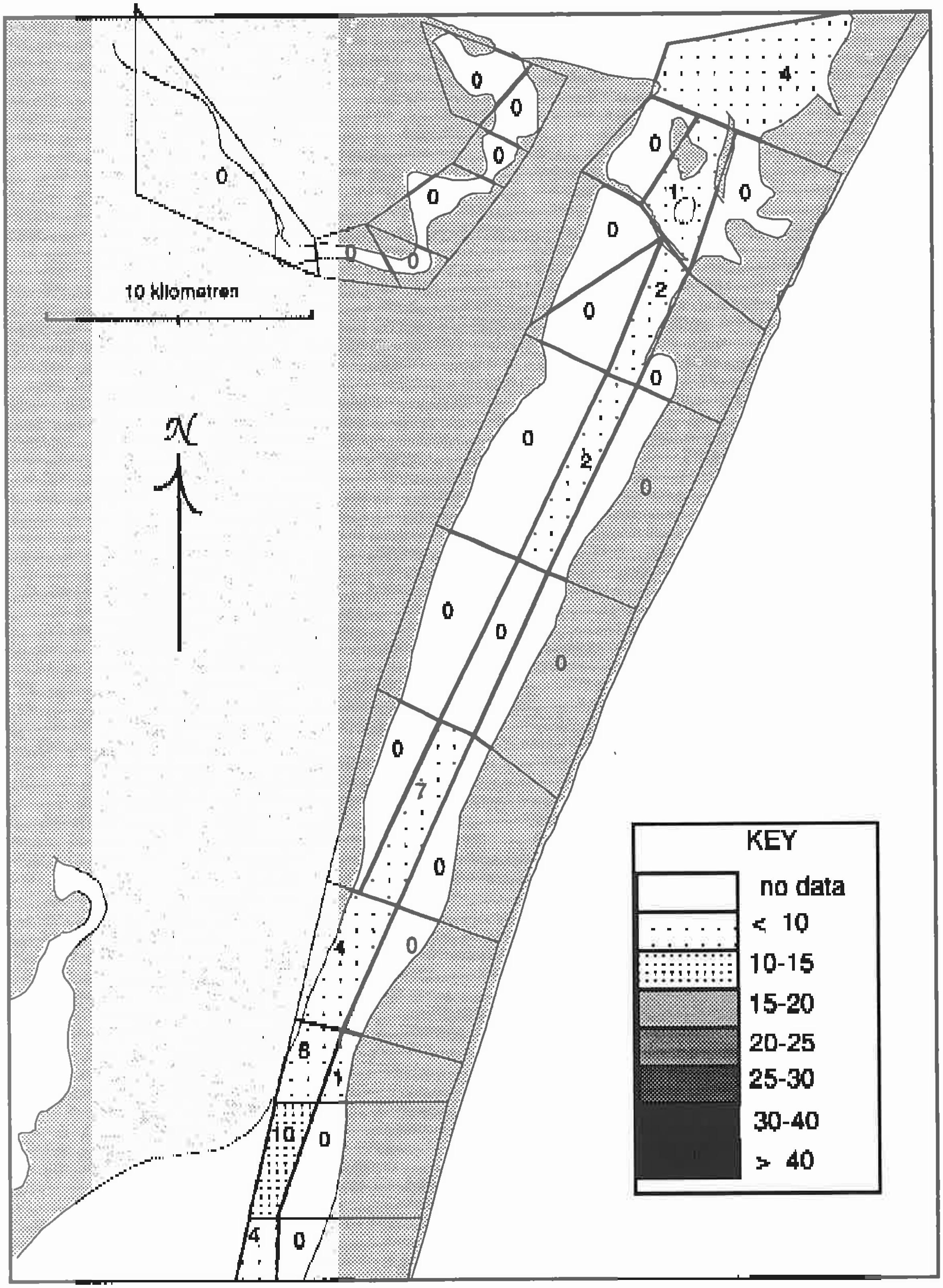


Figure 5-58. Sampling density of WQ-XDDT for Upper Laguna Madre and Oso Bay

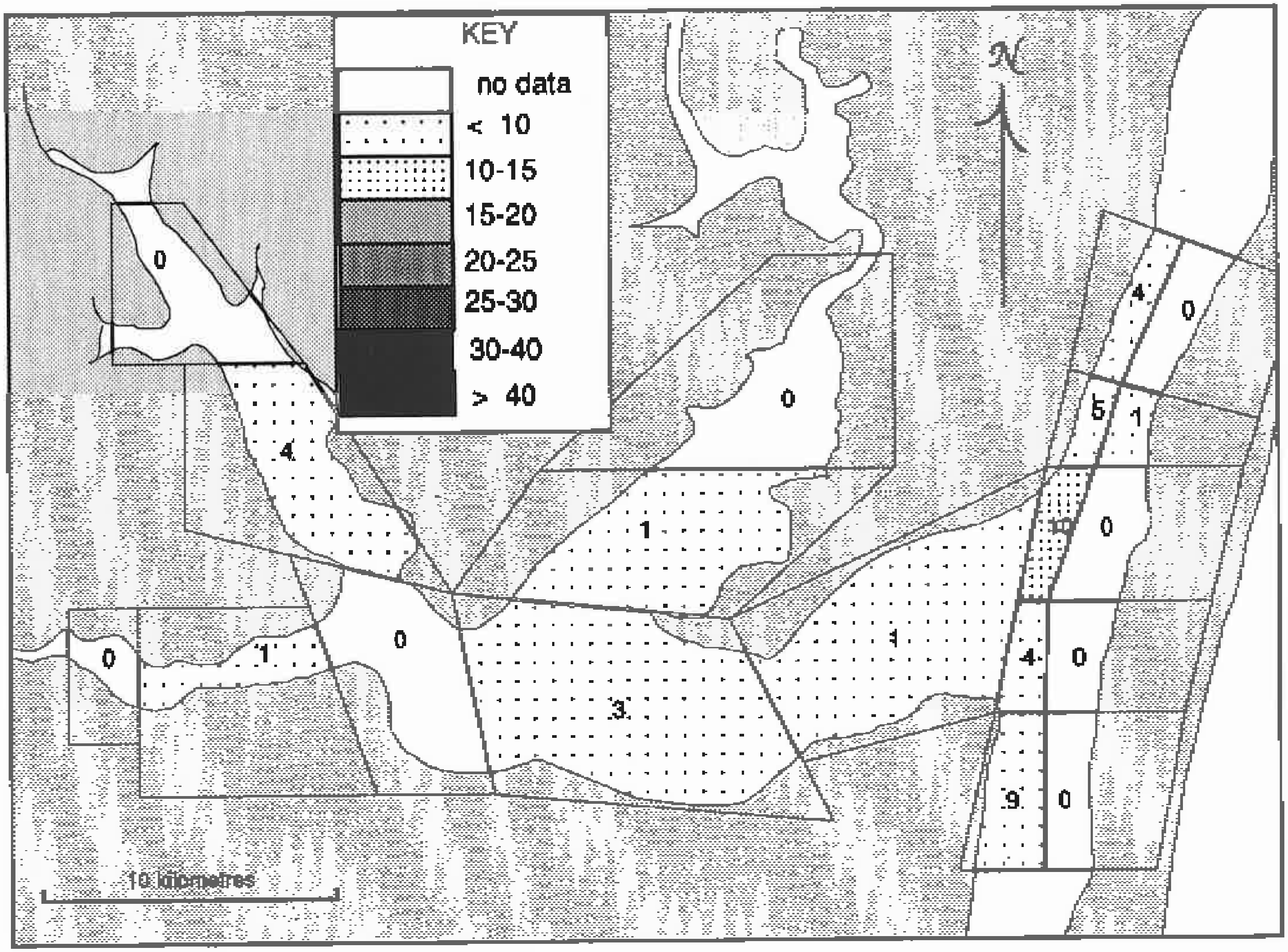


Figure 5-59. Sampling density of WQ-XDDT for Baffin Bay region

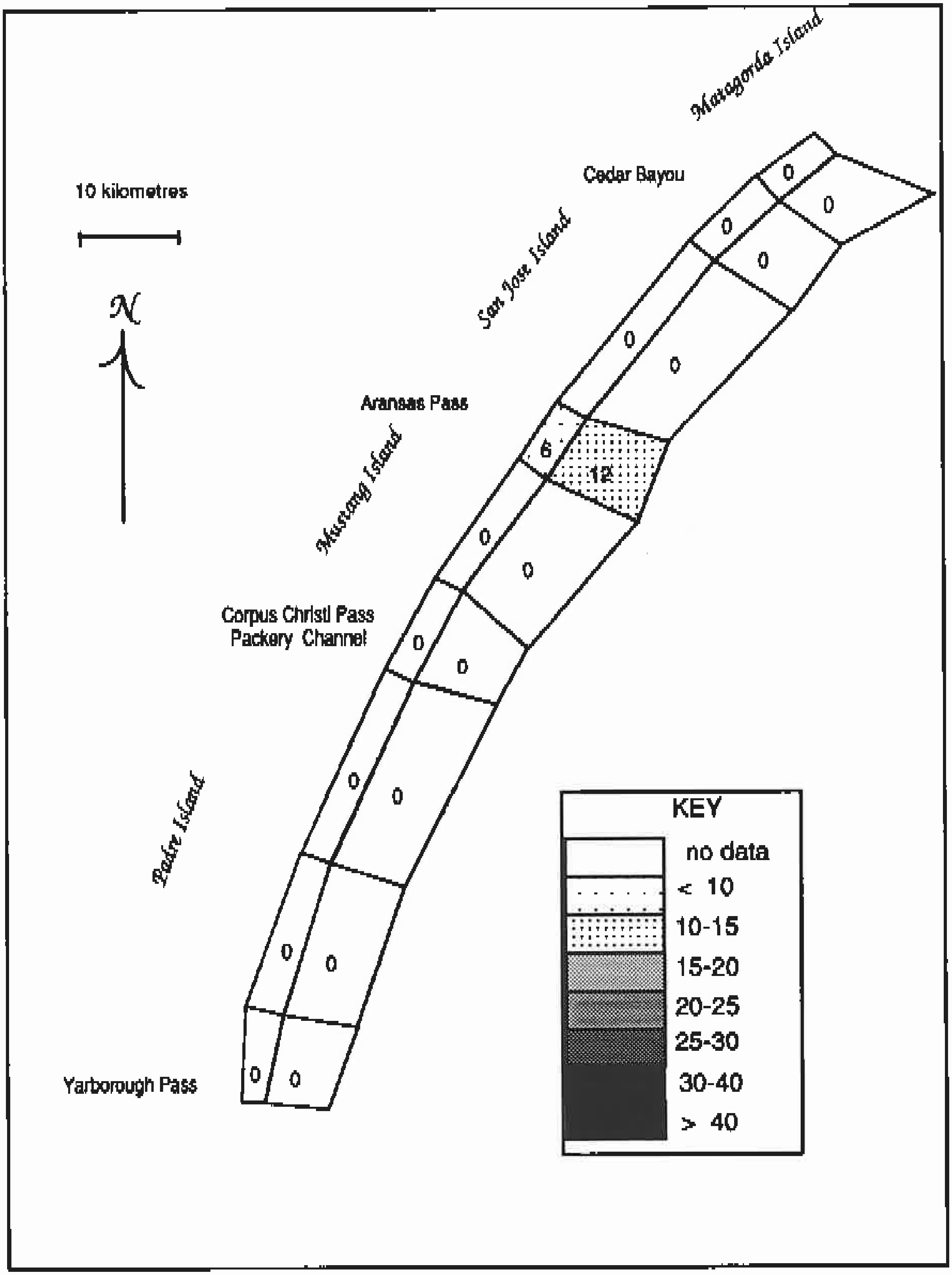


Figure 5-60. Sampling density of WQ-XDDT for Gulf of Mexico

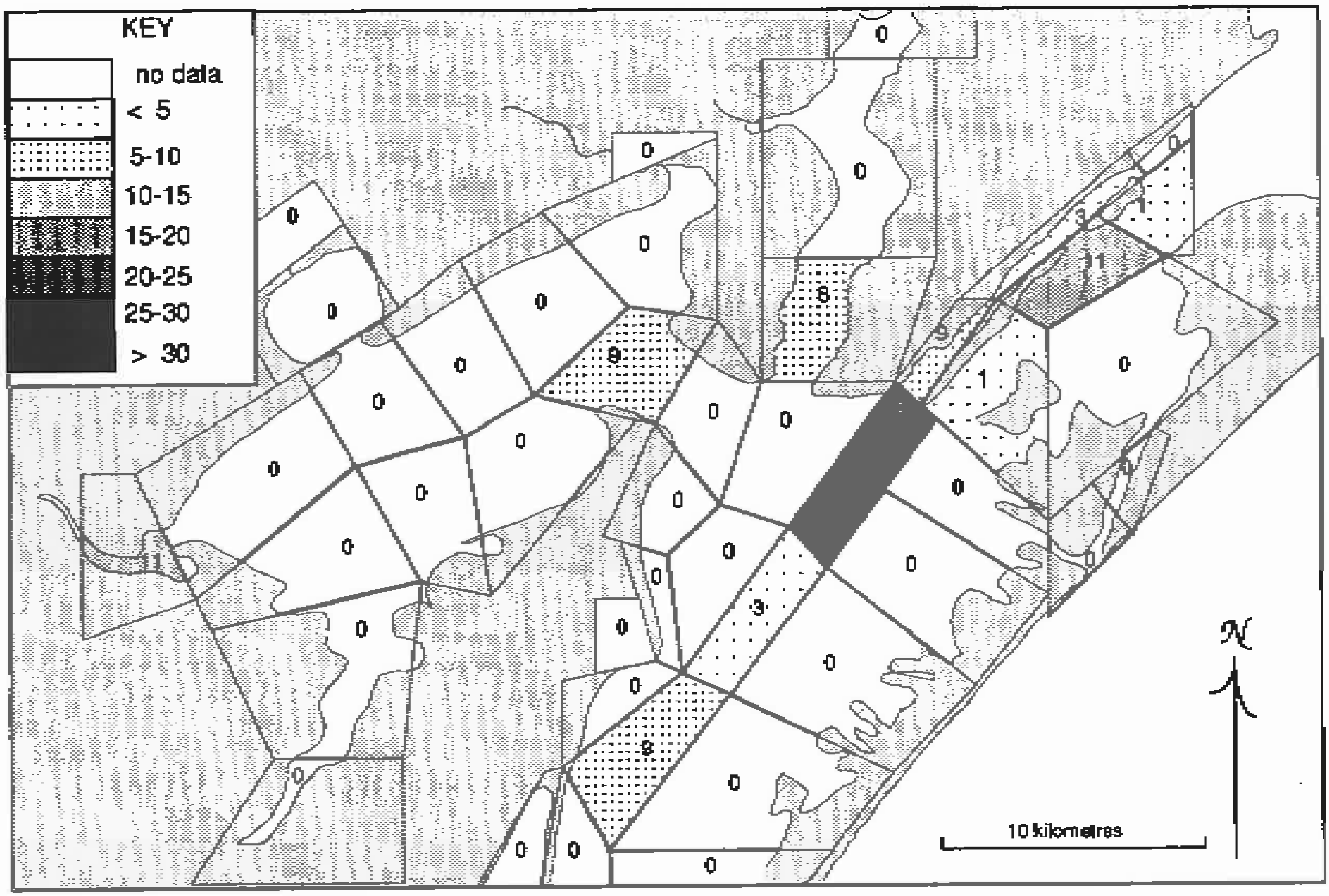


Figure 5-61. Sampling density of SEDO&G for Aransas-Copano system

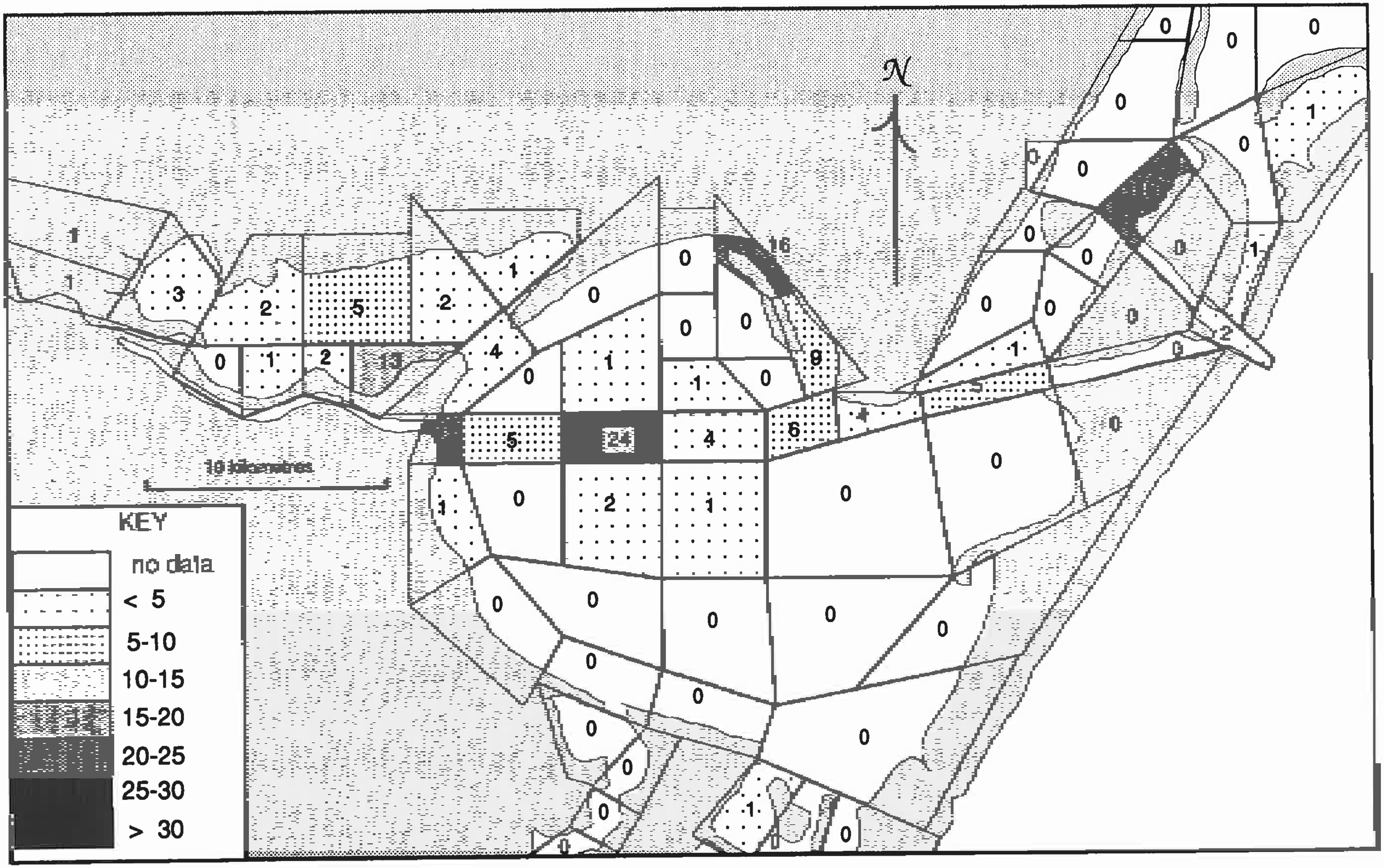


Figure 5-62. Sampling density of SEDO&G for Corpus Christi system

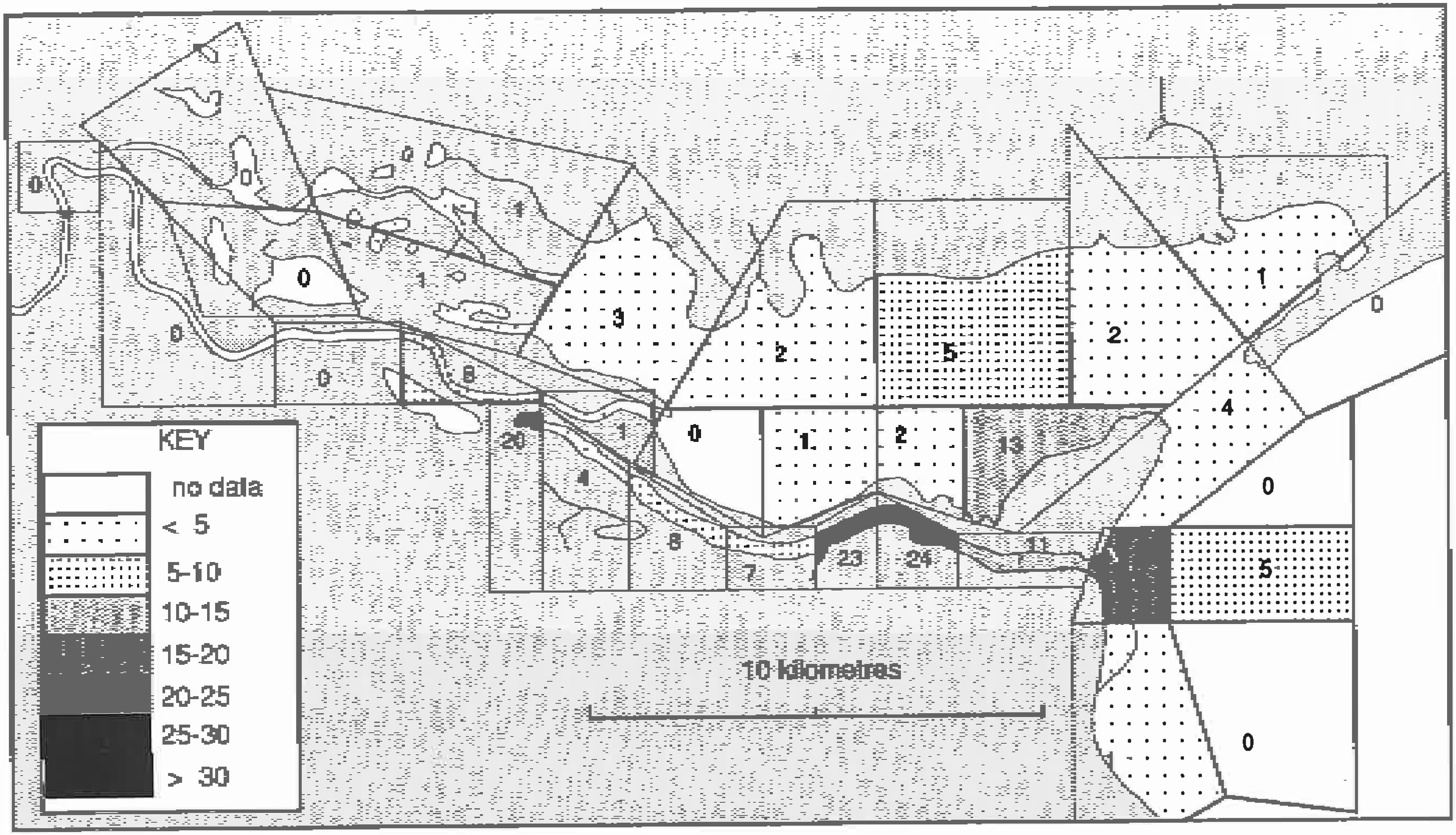


Figure 5-63. Sampling density of SEDO&G for Nueces Bay region, including Inner Harbor

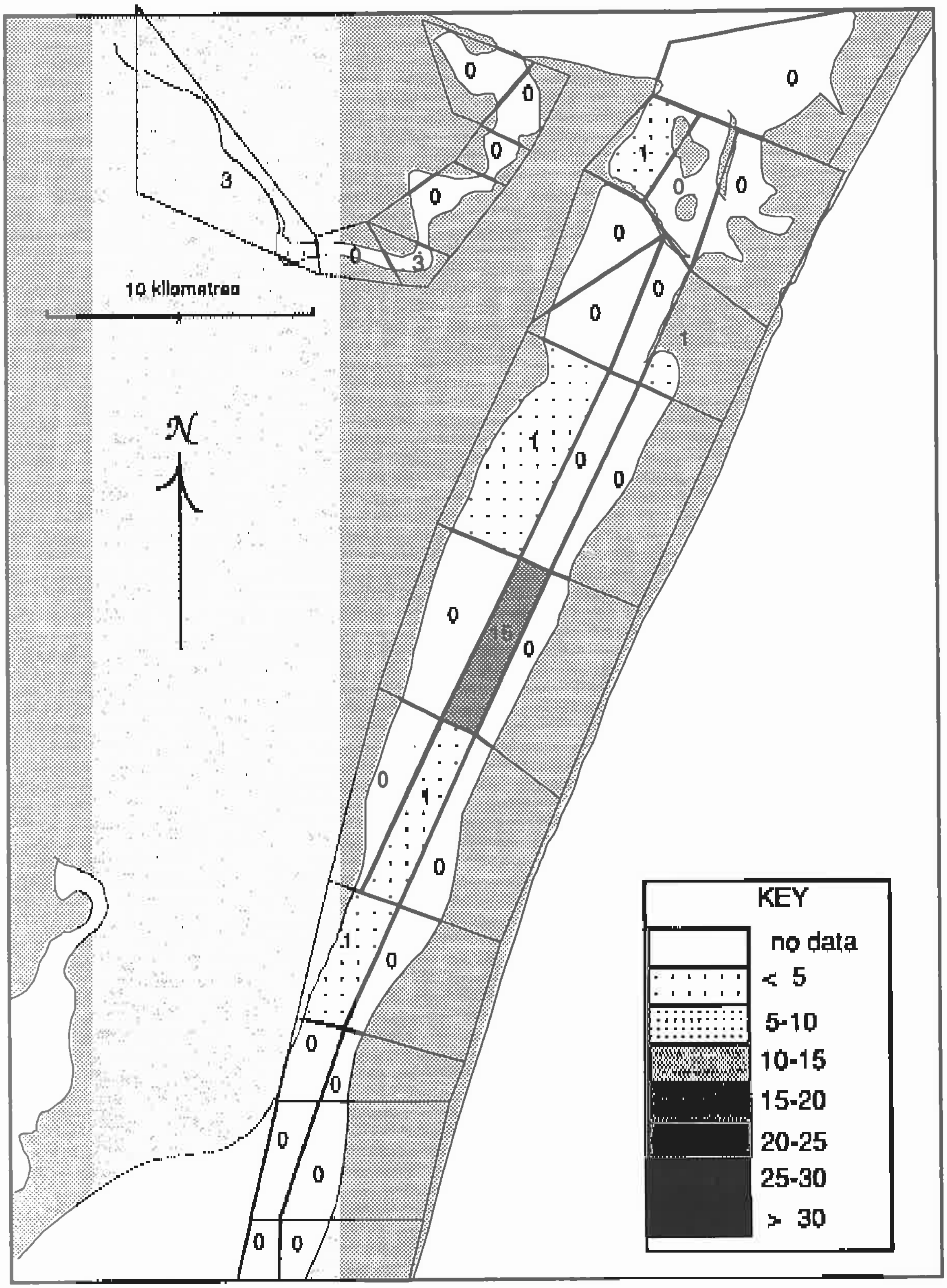


Figure 5-64. Sampling density of SEDO&G for Upper Laguna Madre and Oso Bay

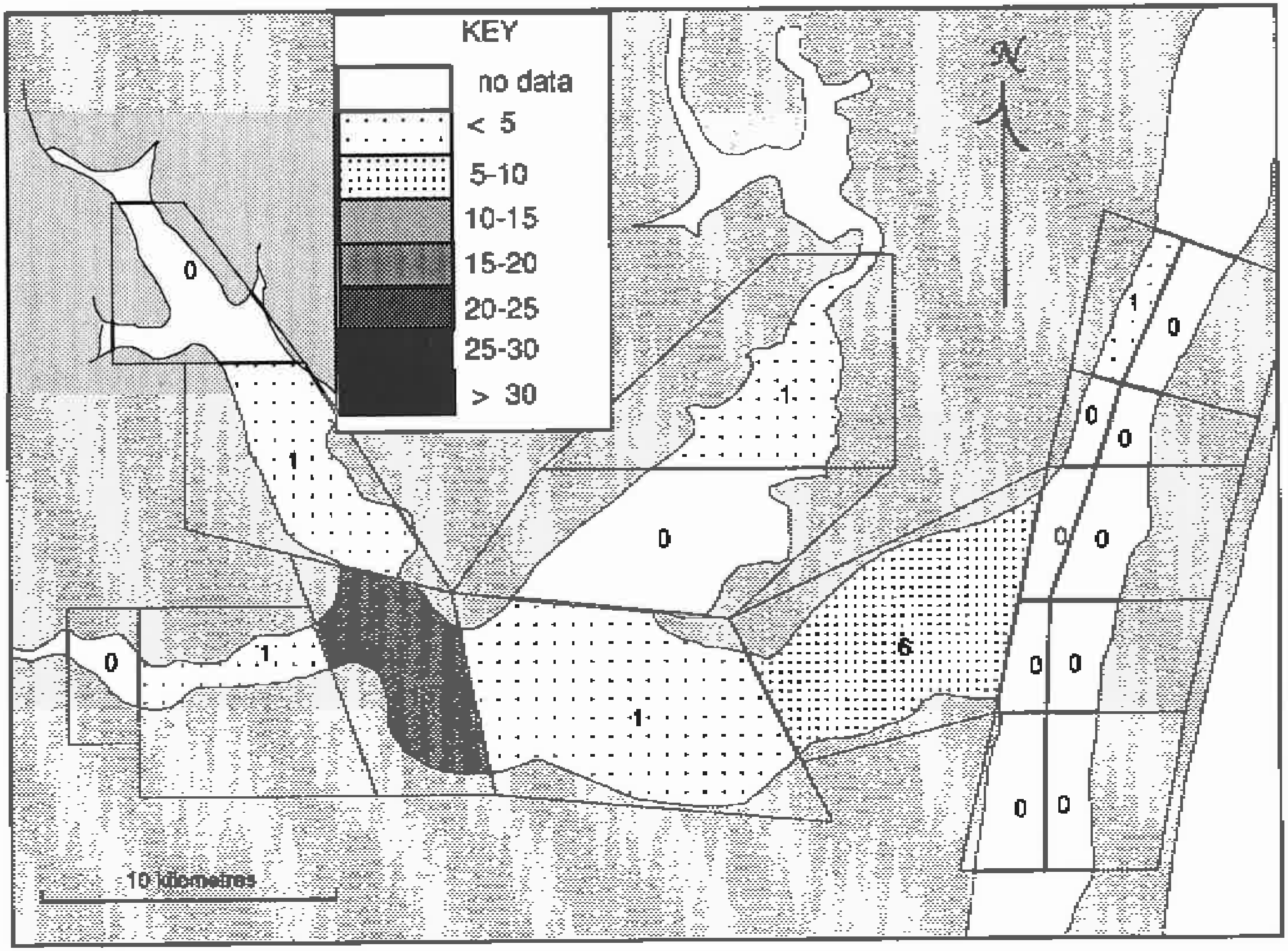


Figure 5-65. Sampling density of SEDO&G for Baffin Bay region

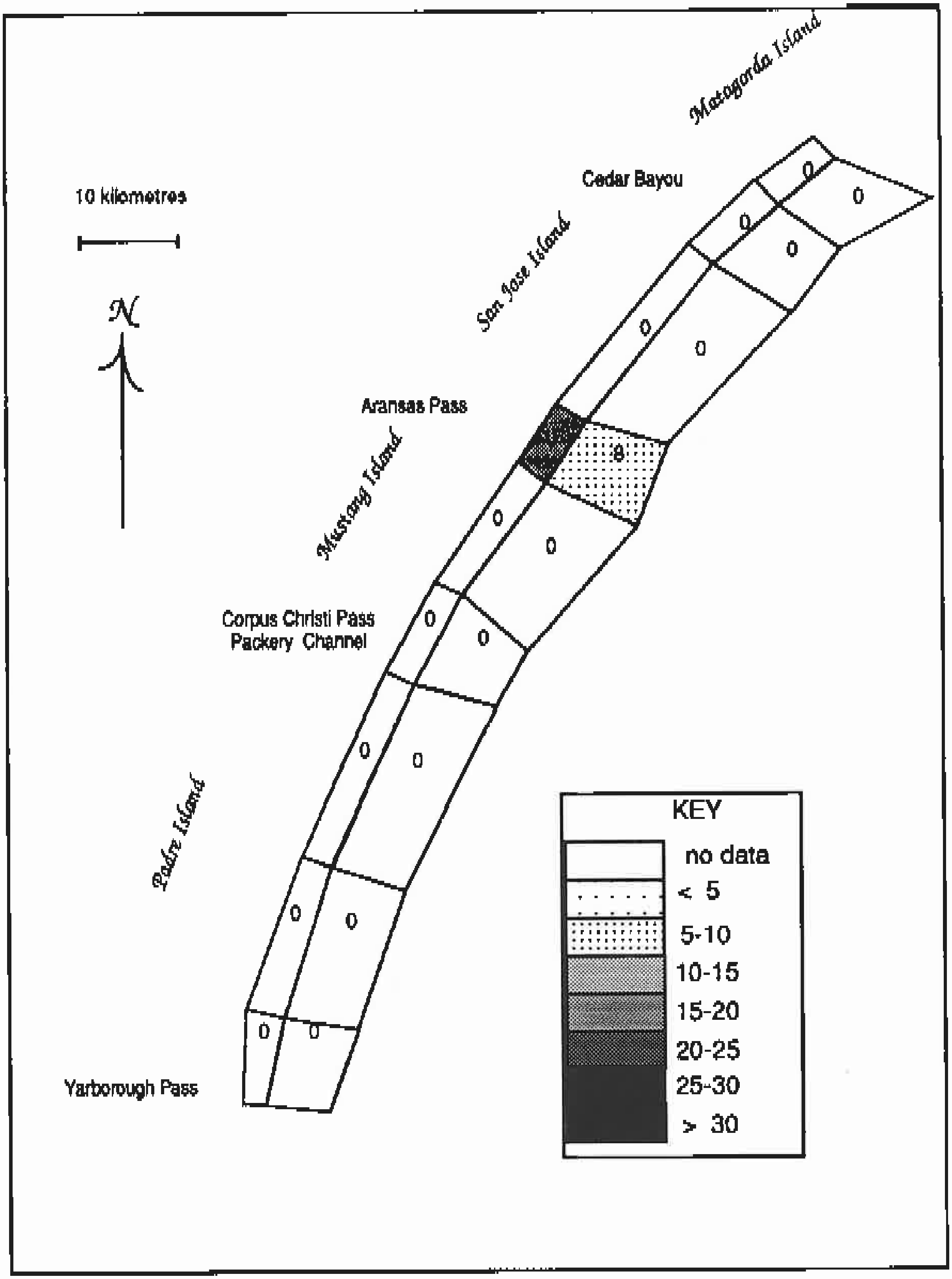


Figure 5-66. Sampling density of SEDO&G for Gulf of Mexico

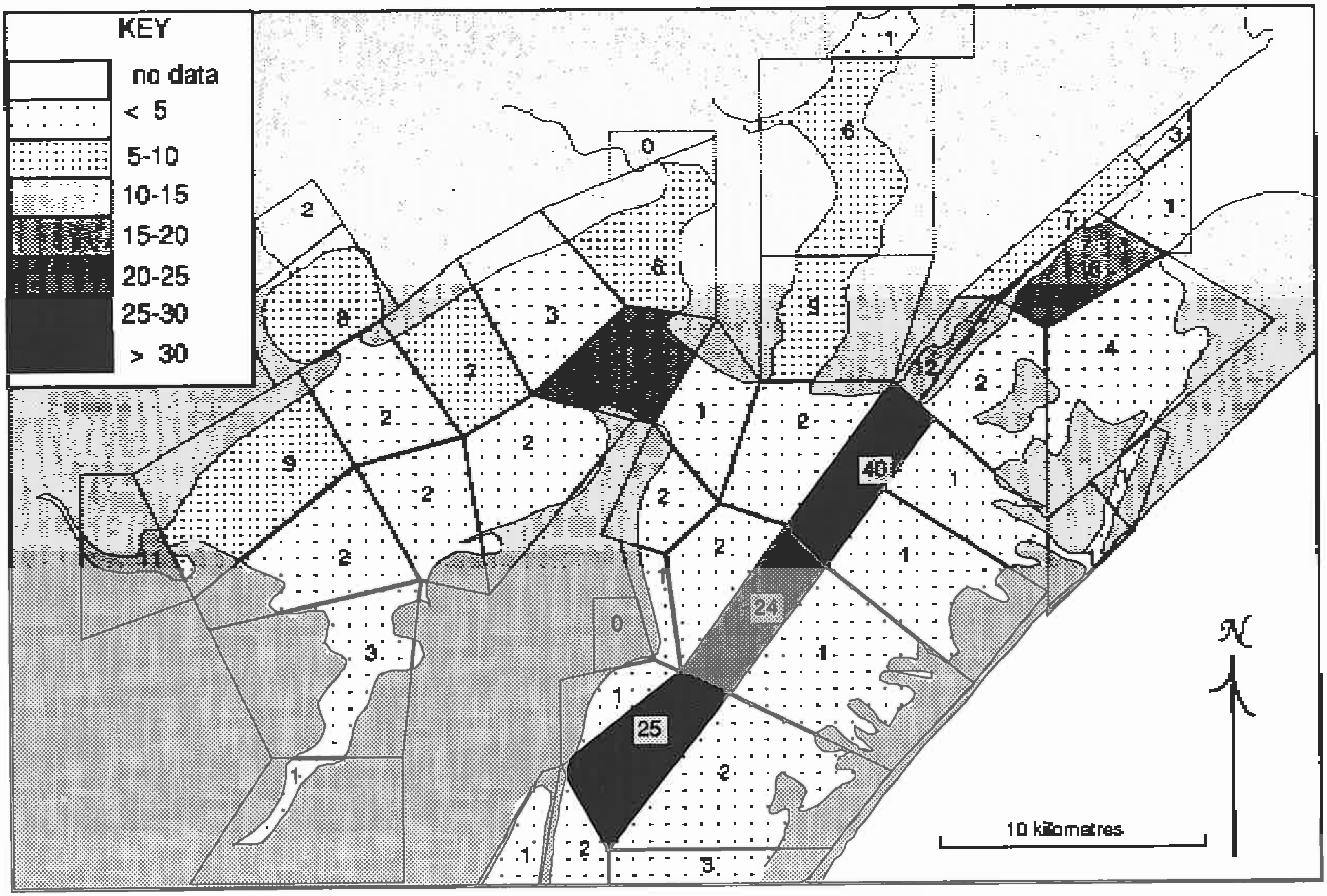


Figure 5-67. Sampling density of SEDMETZN for Aransas-Copano system

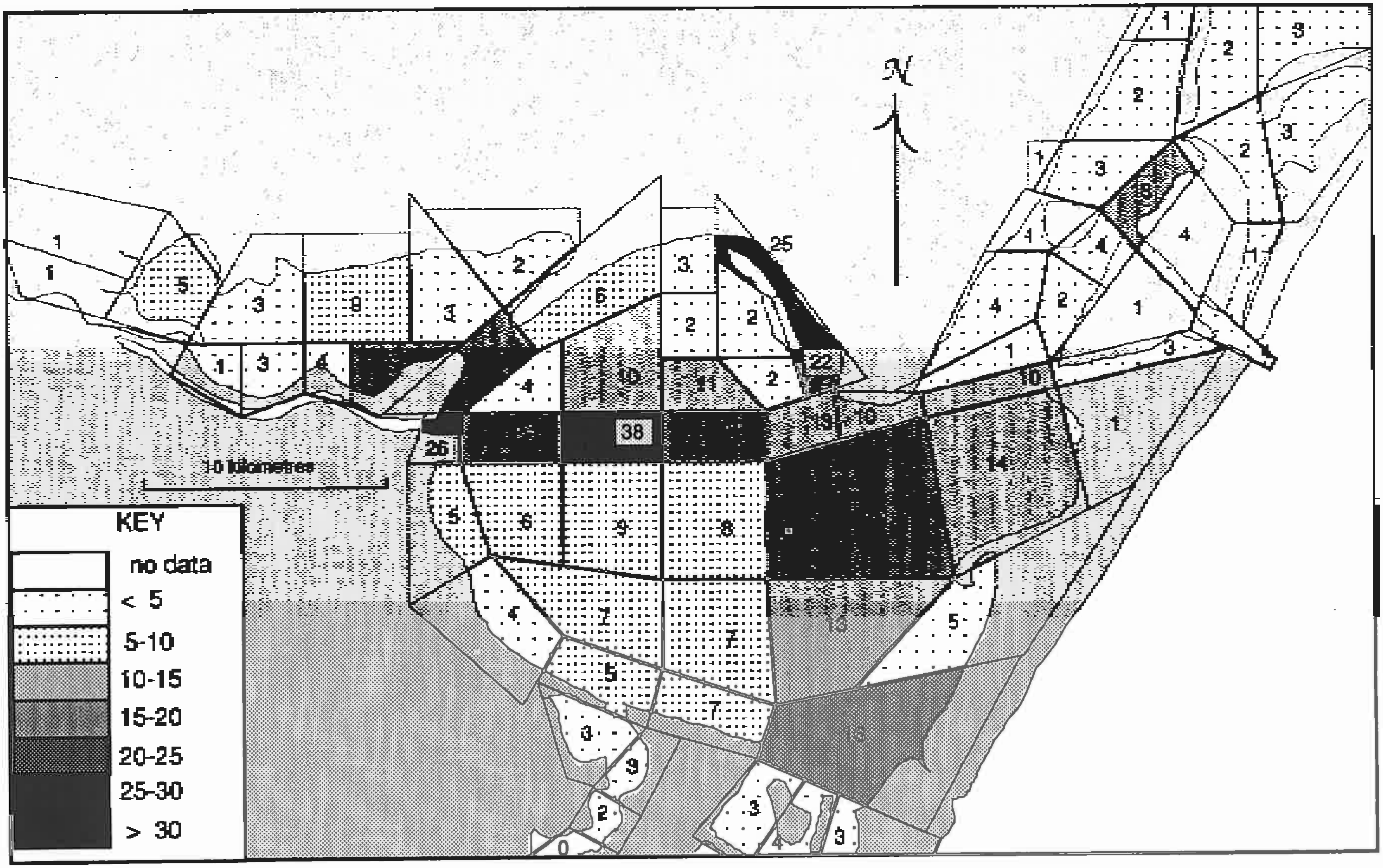


Figure 5-68. Sampling density of SEDMETZN for Corpus Christi system

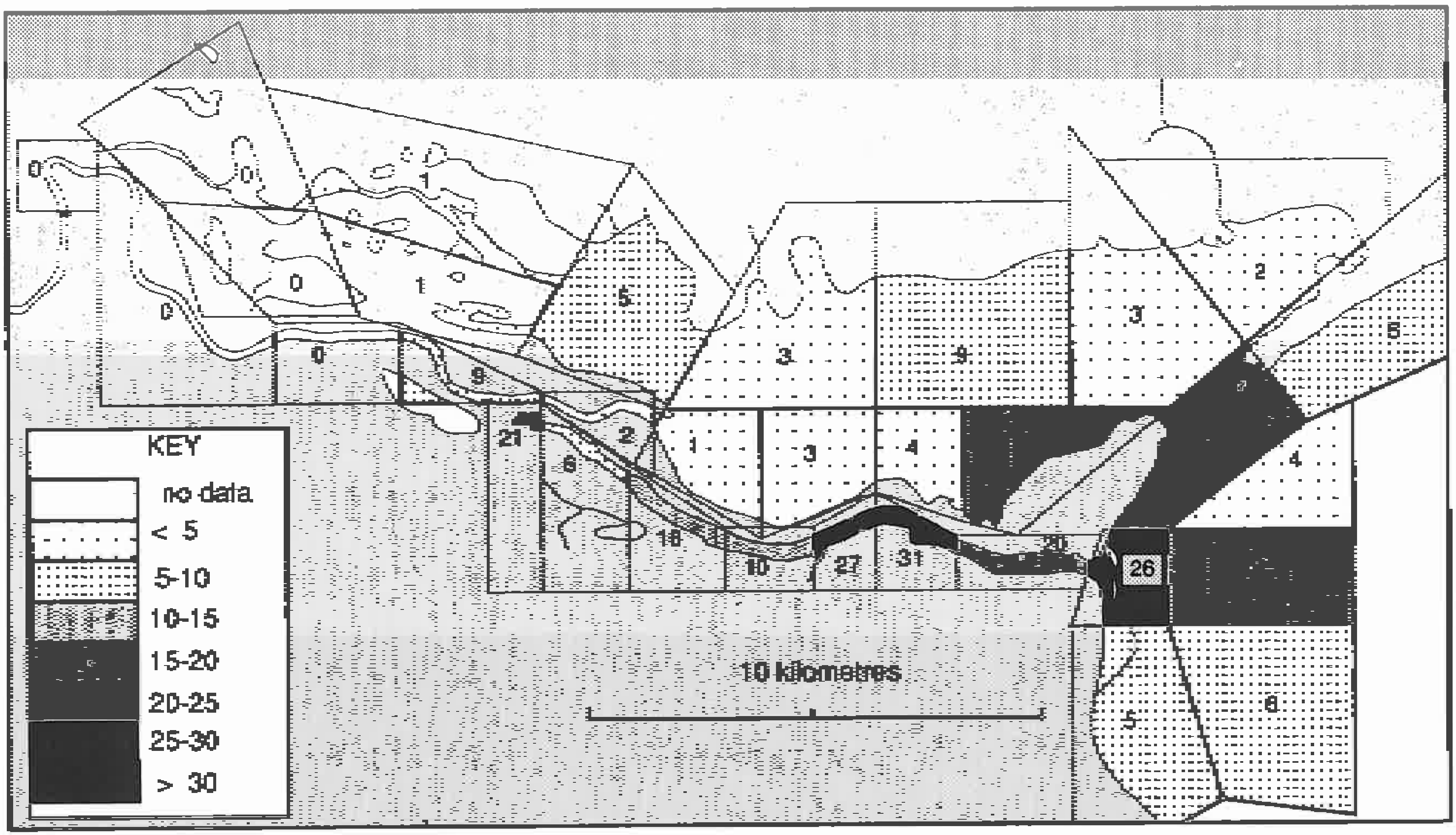


Figure 5-69. Sampling density of SEDMETZN for Nueces Bay region, including Inner Harbor

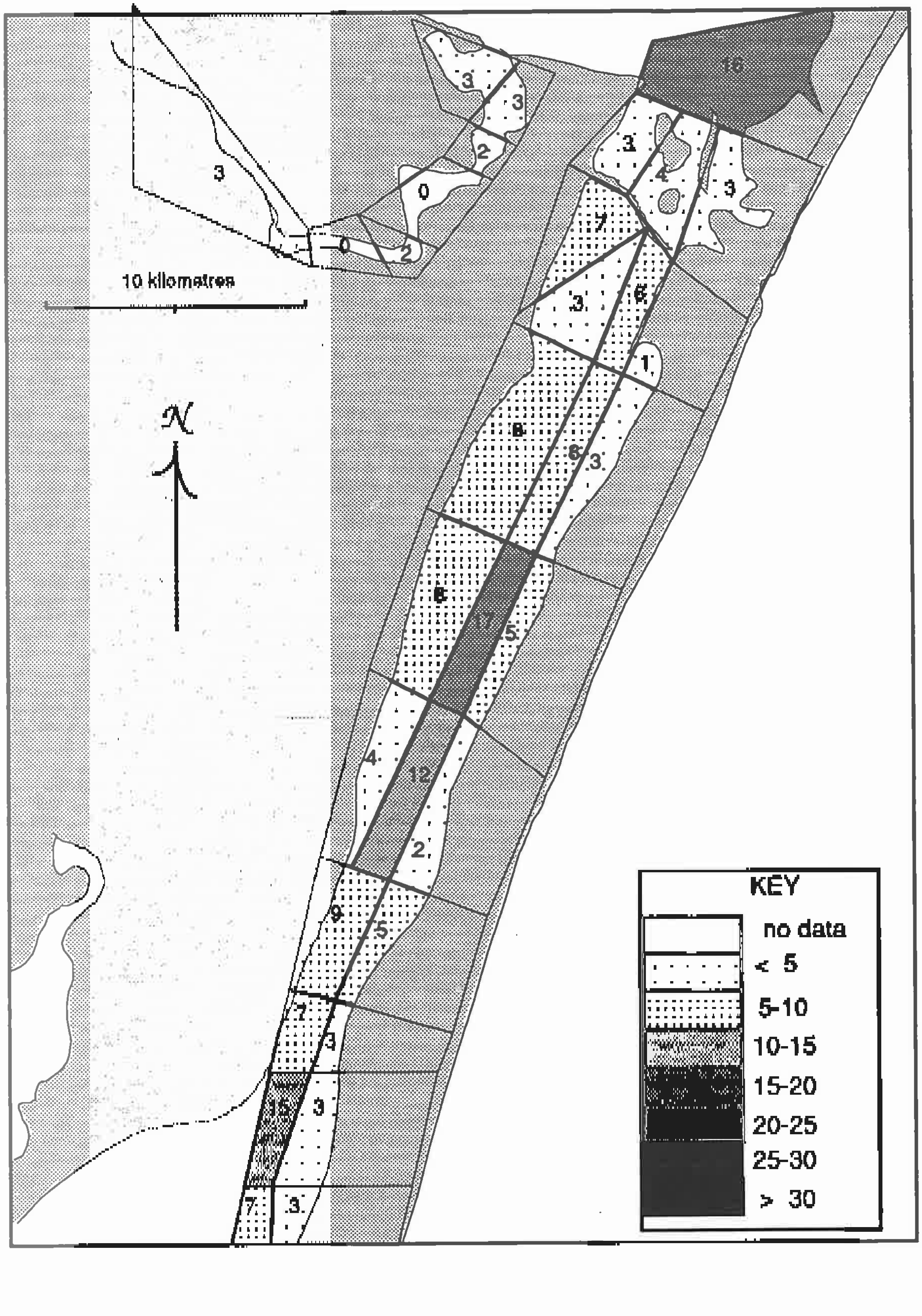


Figure 5-70. Sampling density of SEDMETZN for Upper Laguna Madre and Oso Bay

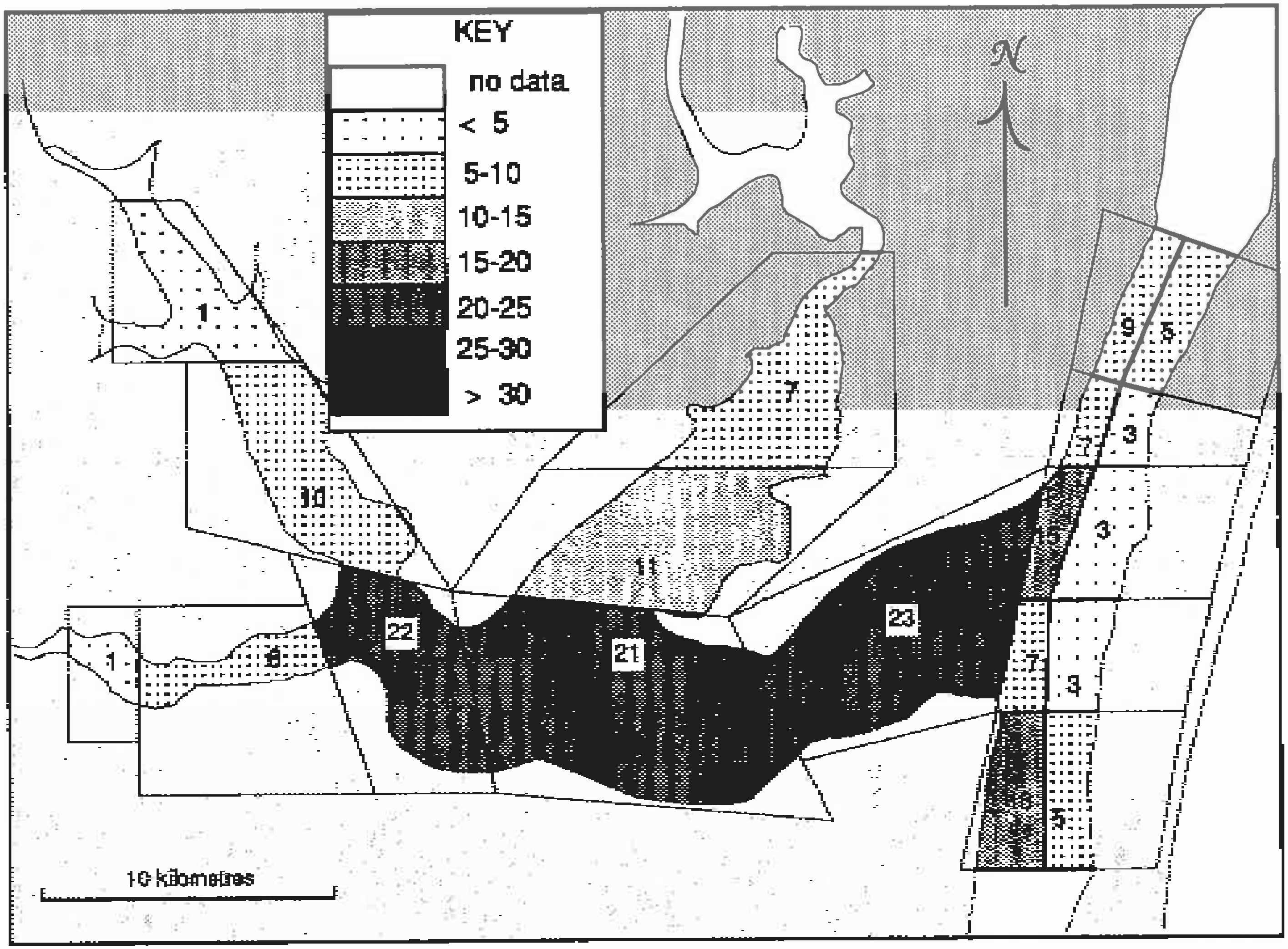


Figure 5-71. Sampling density of SEDMETZN for Baffin Bay region

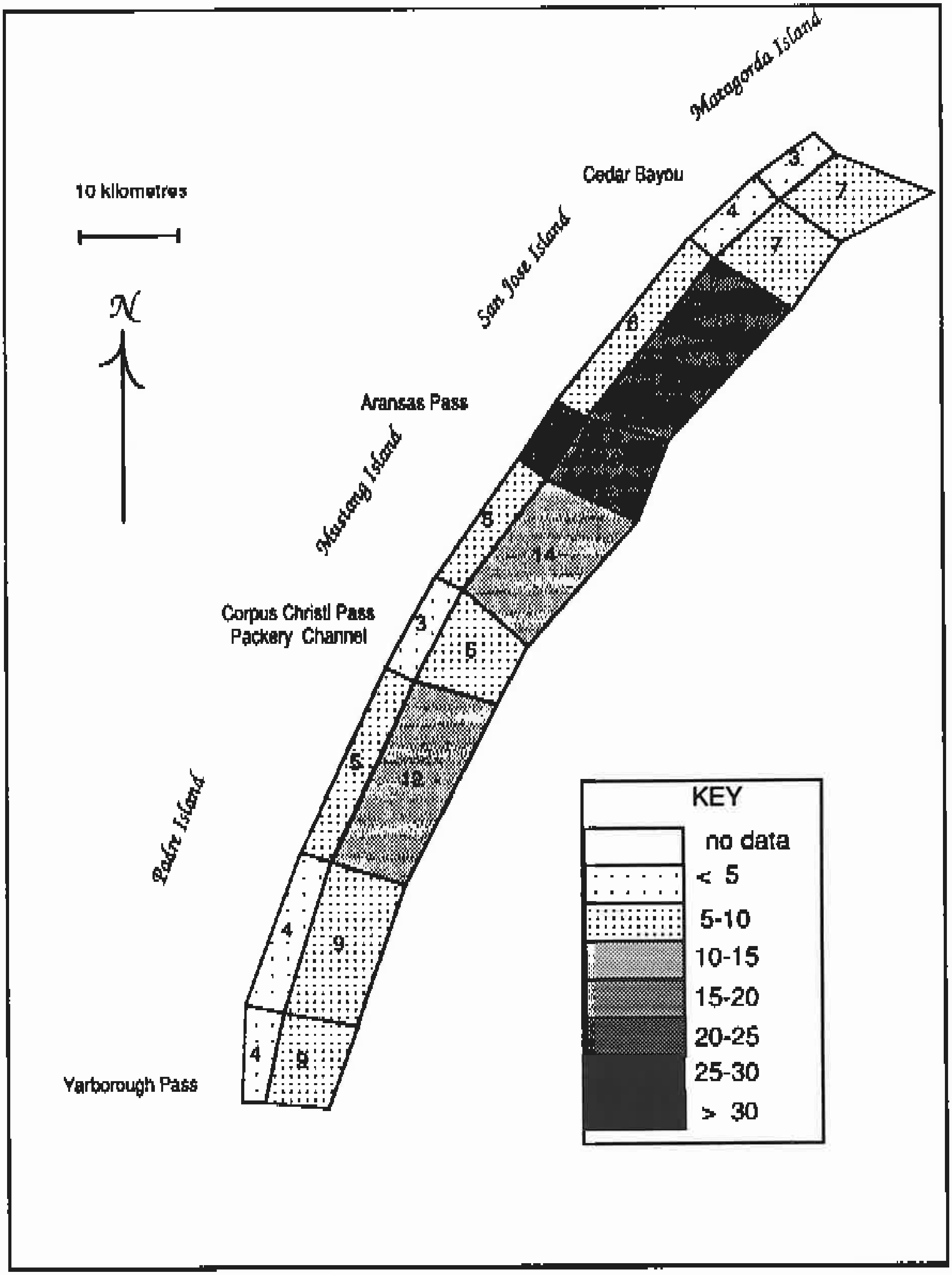


Figure 5-72. Sampling density of SEDMETZN for Gulf of Mexico

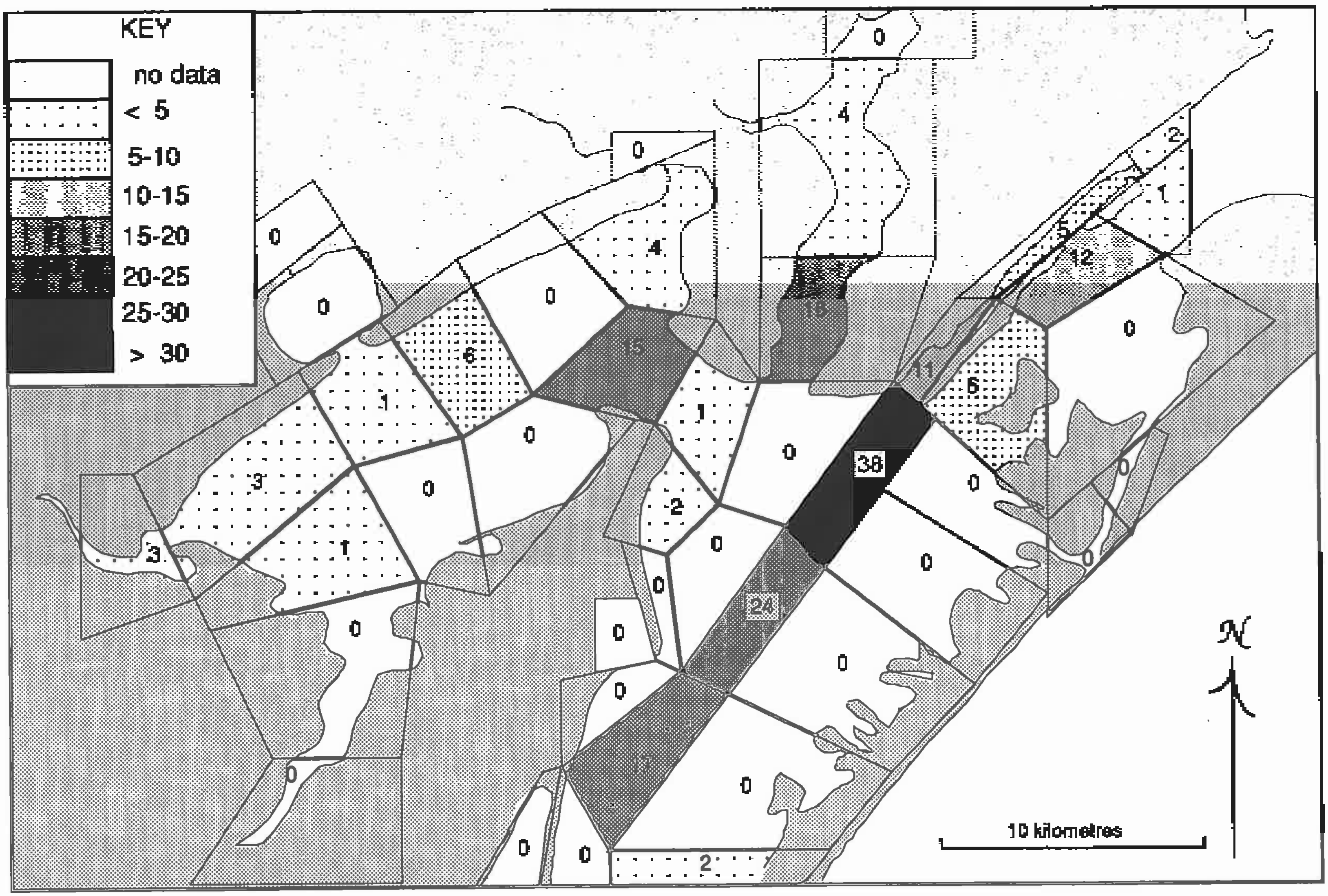


Figure 5-73. Sampling density of SED-XDDT for Aransas-Copano system

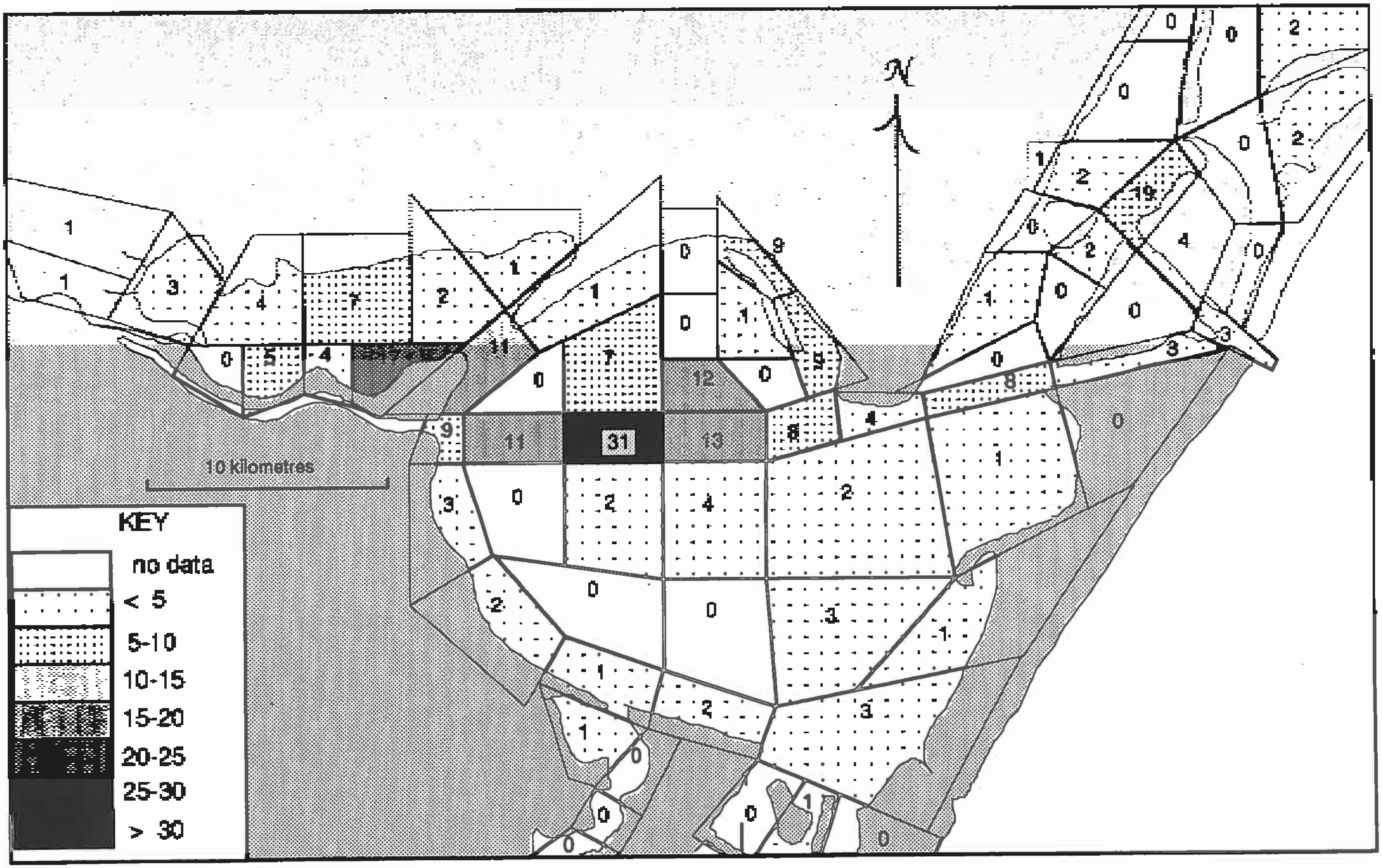


Figure 5-74. Sampling density of SED-XDDT for Corpus Christi system

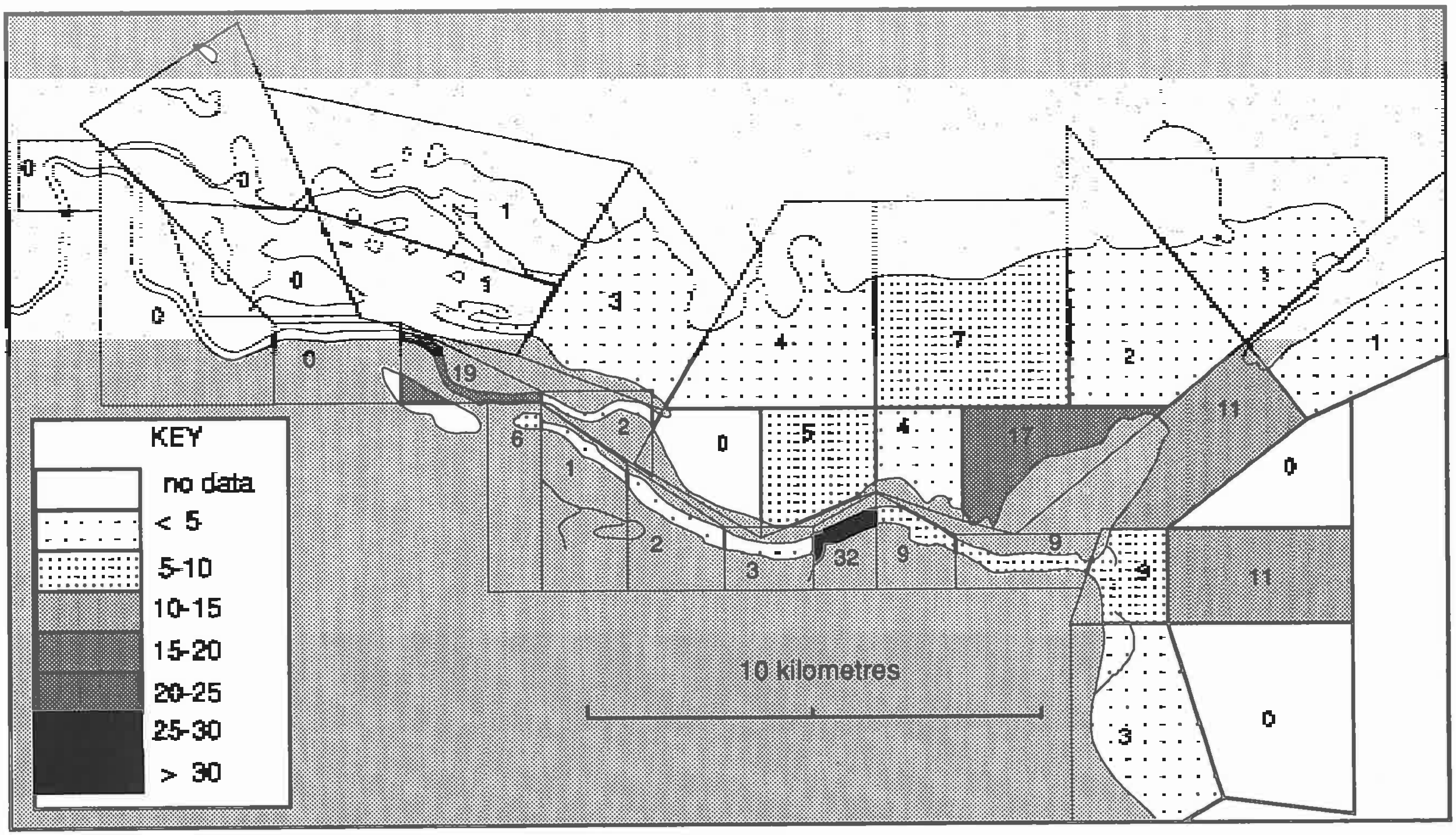


Figure 5-75. Sampling density of SED-XDDT for Nueces Bay region, including Inner Harbor

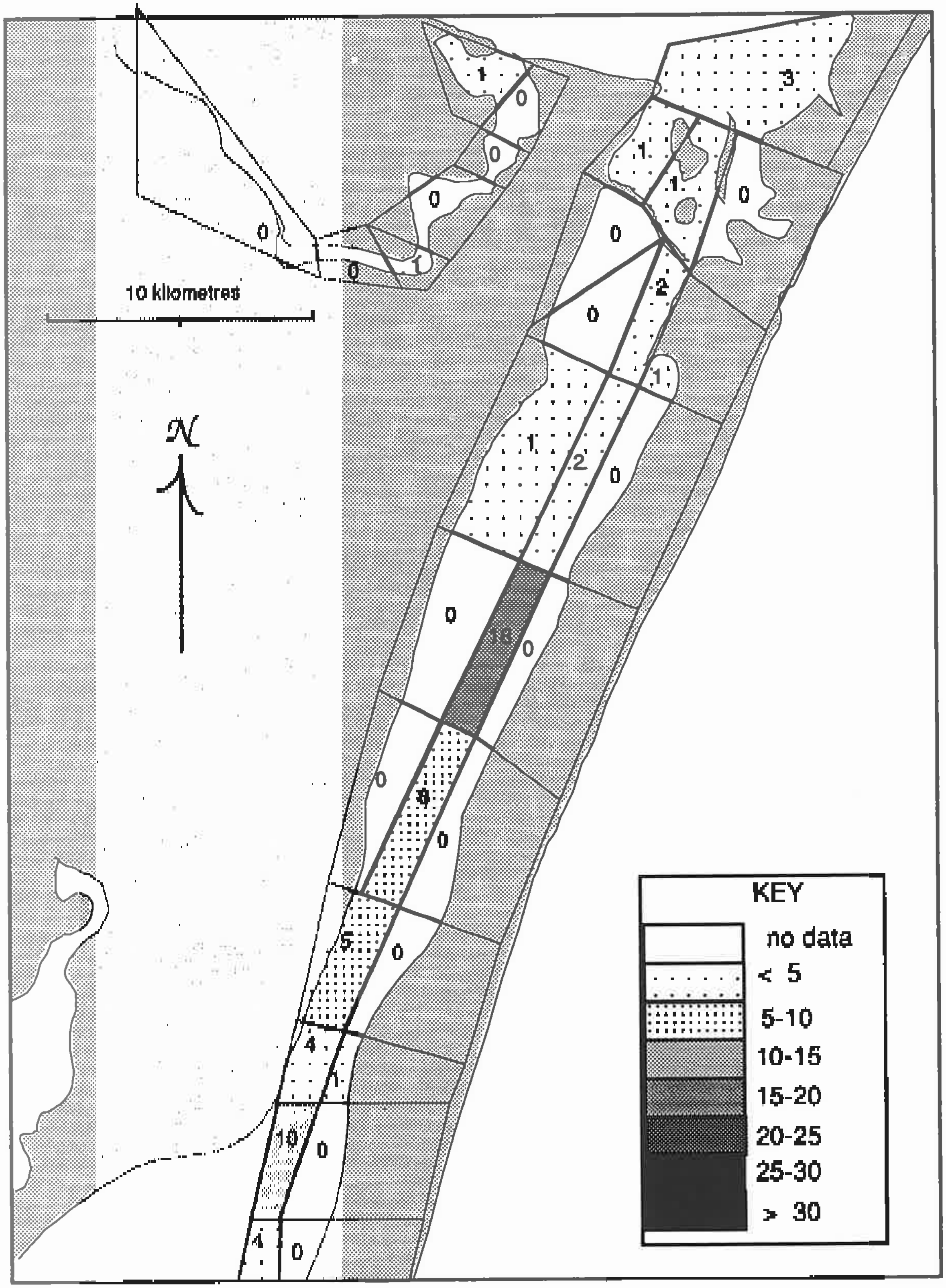


Figure 5-76. Sampling density of SED-XDDT for Upper Laguna Madre and Oso Bay

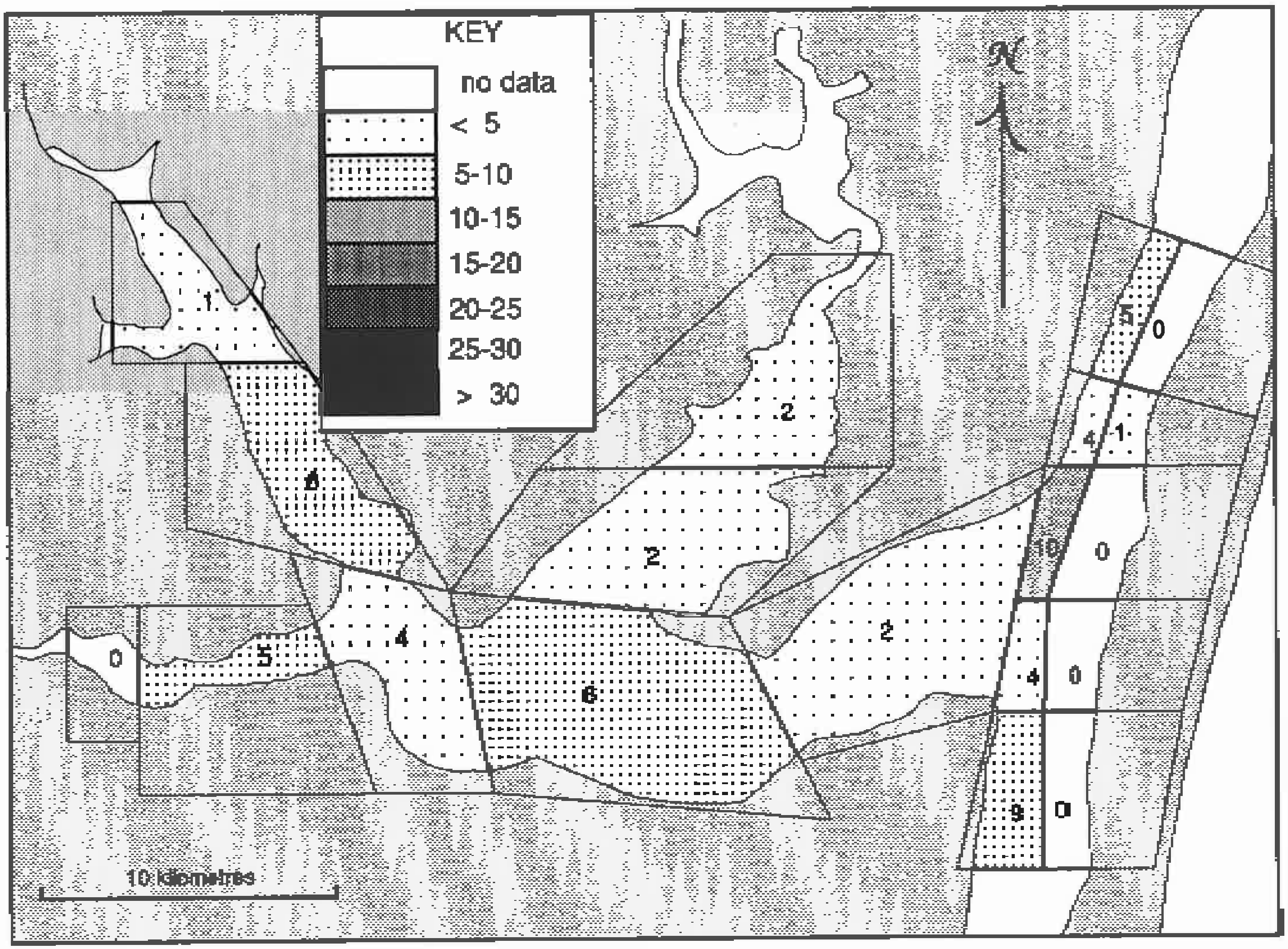


Figure 5-77. Sampling density of SED-XDDT for Baffin Bay region

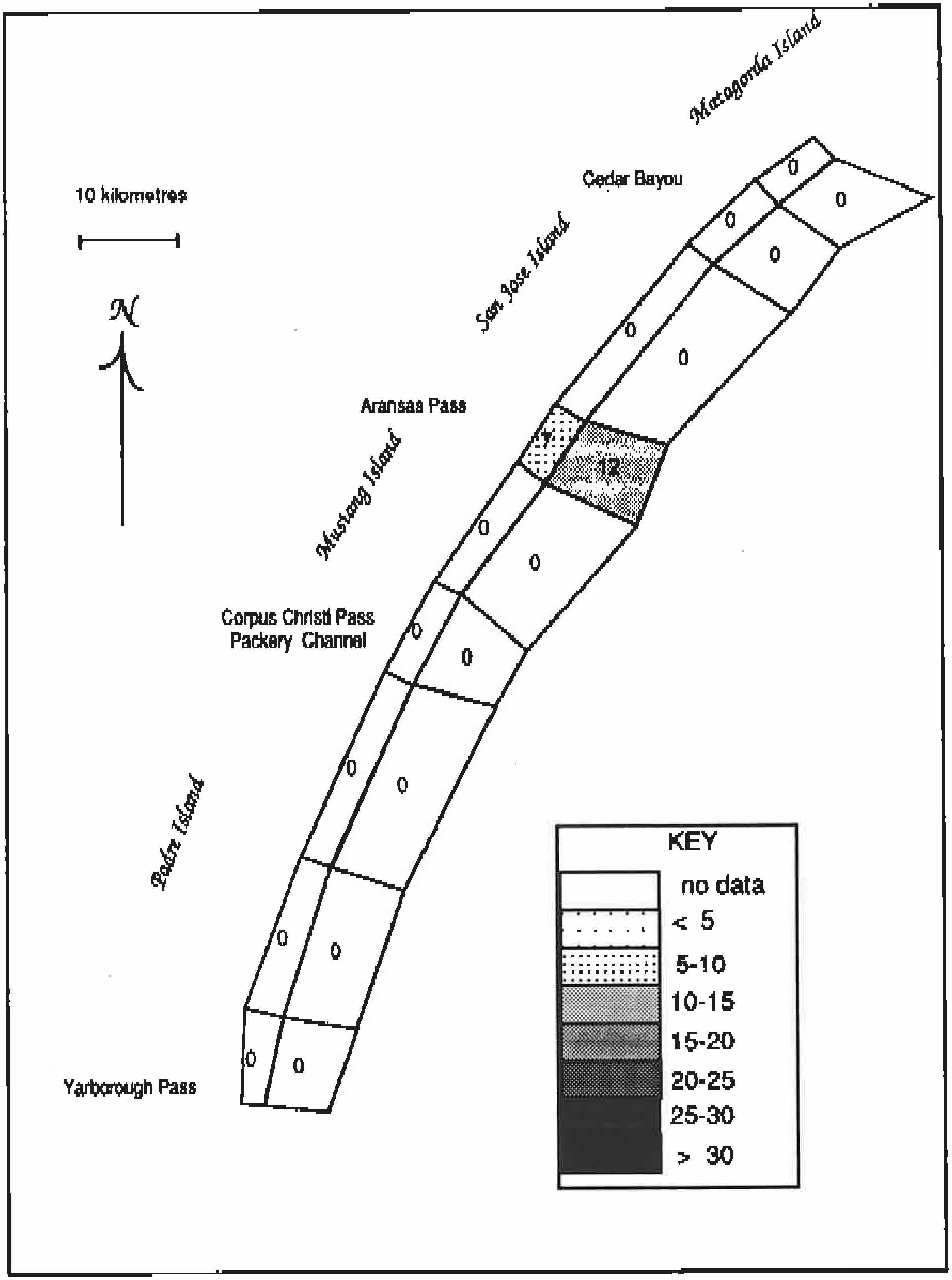


Figure 5-78. Sampling density of SED-XDDT for Gulf of Mexico

that should be theoretically available. Much of the historical data has been lost. Moreover, too many entries in the data record had to be excluded from the analyses presented here because the data were unreliable. It is our belief that this unreliability was not introduced in the original measurement but in the subsequent handling of the data. All of these problems are manifestations of deficiencies in data base management. We differentiate between the management of modern data bases, i.e. more-or-less contemporary data streams, discussed in Section 5.3.2, and the management of historical data bases, treated in Section 5.3.3.

Recommendations for both collection and management of data from Corpus Christi Bay are presented in Chapter 10.

5.3.1 Deficiencies in sampling strategy and data distribution

Because Corpus Christi Bay is a complex estuarine system, spatial gradients in concentration are the norm, created by circulation, intermixing of fresh water and seawater, and sources of pollutants. Positioning information is indispensable to properly interpreting measurements, the requisite accuracy being dictated by the region of the bay and the prevailing spatial gradients. Despite this, positioning information associated with water samples is frequently of inferior quality, and may be so erroneous that the data is unusable. For most sampling programs, positioning is a three-step process, each step of which is subject to errors:

(1) bringing the boat or sampling platform to its desired position
(2) rendering the sampling location on a map
(3) determining the geographical coordinates

For stations in a confined waterway with shore-based landmarks, or for stations on a well-marked navigation channel, positioning of the platform (1) can be carried out accurately in the field. When these landmarks, e.g. bridge structures or navigation aids, are *also* accurately located on a published map, Step (2) of rendering the locations on the map can be performed with equal accuracy. In the open waters of the bay, positioning in the field is more problematical, and placing the occupied position on a map is just as questionable. The traditional method of "eye-balling" is often used, relying upon distant landmarks, attempts to hold constant headings, estimating distance by speed (itself a judgment) and travel time, and similar exercises. The potential for error is considerable. The BEG (Project 12) and USF&WS (Project 27) both used this method, as did SWRI, USGS, TNRCC and many others.

Many of the data sources used in this study carried out only the first two steps. This project then performed the third step of reading the geographical coordinates from the map; the process and Q/A procedures are described in Sections 4.1 and 4.2.2. Some data sources, such as the Texas Parks and Wildlife Department (Project Code 3) and U.S. Geological Survey (Project Code 11), carried out Step (3) and provided only the final latitude/longitude coordinates. We of course have no means of verifying these positions. Some of the TPWD stations plotted out in impossible positions, such as off North Africa, and one of the USGS stations plotted out in the middle of downtown Corpus Christi. Data from these stations had to be excluded from the compilation. But these raise the question of how many other stations are also mislocated, at positions that happen to lie within the study area so that the error cannot be detected.

In recent years, technology that would appear to eliminate all of these problems has become inexpensive enough that it is beginning to be used in sampling programs, *viz.* radionavigation systems, primarily LORAN, and geographical positioning systems (GPS). These systems include an onboard processor that provides immediate feedback to the operator, and effectively collapse all three of the above steps literally into a push of a button. It seems almost too good to be true.

LORAN (nee LORAN-C, before the original LORAN system was dismantled about a decade ago) is a hyperbolic radioposition system that operates in the LF band. It is subject to two primary sources of error: poor geometry (i.e., oblique crossing angles of the LORAN lines of position) and propagation noise, to which LF signals are particularly prone due to the variability of the ionospheric waveguide. Nominal accuracy cannot be expected to be better than 0.5 km, and may be 2-5 times this under poor atmospheric conditions (which are not obvious to the user). The internal processor for the onboard receiver does not provide the user with any measures of uncertainty, but produces latitude/ longitude coordinates to a meretricious 5 or 6 significant figures. LORAN-derived coordinates were provided for the MSI 1988-89 stations in Corpus and Nueces Bays (Project Code 17). In comparison to a hand-plotted map, these looked "about right" but upon closer verification, 12 out of 33 had to be mapped and their coordinates determined by hand because the LORAN positions were over a kilometre in error. One of these, in lower Nueces Bay, plotted out according to LORAN in a cow pasture near Odem. Another, in the Ship Channel near Harbor Island, plotted out in the surf off of Packery Channel.

GPS does not appear to be much more reliable, the vaunting of vendors notwithstanding. The signal is subject to error due to propagation between satellite and receiver. To realize the advertised accuracy, the receiver/processor must remain stationary for some time to integrate out the noise in the signal, which is often difficult to manage in a boat. ("Differential mode" using a fixed base station signal can considerably improve the performance.) Again, the black-box processor produces latitude/longitude coordinates to an impressive precision. The stations for one of the Laguna Madre data sets used in this compilation were determined by GPS: they plotted out 10 km in the Gulf of Mexico. (To be fair, though, it must be stated that the error in these positions may have been introduced in post-processing, and not from the GPS unit itself. We have too little information about the actual field and data-processing procedure to judge.)

Geographical distribution of sampling stations is frequently driven by the management objective of the agency, appropriately; but this also means that the data may not be placed most strategically for water-quality characterization. For example, Nueces County Health Department monitors to ensure the quality of water for contact recreation, and therefore places its stations in areas where such recreation is concentrated, which happens to be mainly along the shoreline adjacent to the City of Corpus Christi. Similarly, TNRCC monitors pollution mainly in areas expected to be subjected to contaminants, such as the Inner Harbor and La Quinta Channel. Texas Department of Health monitors areas subject to shellfish harvesting. The Corps of Engineers concentrates its monitoring activities in the proximity of ship channels. The utility of any of these stations to a characterization of water quality is a matter of fortuity.

If one wishes to best determine properties of a water mass with a single station, it is apparent that one would monitor in the midpoint of the water mass. The reason, perhaps sensed intuitively, is that such placement will ensure that the measurement will be minimally subject to random spatial variation since here the water mass would be expected to be most uniform. In contrast, near the boundary of the water mass, measurements would be much more variable, and therefore less reliable, due to intermixing with waters from adjacent masses. This same principle ideally would be applied in placing sampling stations in an estuary for monitoring data to characterize water quality, i.e. placing the stations where water properties would be expected to be most homogeneous, and avoiding those areas where large spatial gradients would occur. For example, one would monitor near the center of a secondary bay, and would avoid monitoring at the inlet between this bay and the larger connected system. Yet an inspection of the sampling density of the data from Corpus Christi Bay shows that the inlets between systems, such as the entrance to Nueces Bay, the pass between Copano Bay and Aransas Bay, and the Upper Laguna Madre at the JFK Causeway, are most frequently monitored. This is an unfortunate manifestation of the fact that these areas are generally preferred because station location is much easier.

Of course, the optimum sampling distribution would be a dense, uniform distribution throughout the system. This is rarely possible, especially for a monitoring program, in contrast to a one-shot survey, therefore we have followed the strategy of collecting the data into hydrographic areas for spatial analysis. The sampling density figures given in the preceding section evidence a high degree of heterogeneity in sampling. As expected, Corpus Christi Bay *per se* is sampled most often, with the upper and lower systems of Aransas-Copano and Baffin being sampled much less frequently. The sediment samples are particularly poorly distributed, with the outer bays not being sampled for many parameters in years.

As noted in Chapter 1, the time history and temporal intensity of data collection in the bay are of central importance in assessing the data base, from two standpoints. First, a long period of data collection is necessary to encompass the range of variation of external conditions to which the bay is subject and which contribute to the variance in the parameters. Second, from the standpoint of determining time trends, a long period of record is absolutely indispensable. Moreover, a long period of record is much more important in improving statistical reliability than a reduction in the noise in the data.

The time history of sampling in the Corpus Christi Bay system since 1955 is roughly indicated in Figs. 5-79 through 5-89. These show total measurements per year in the entire study area including the adjacent Gulf of Mexico for the indicated parameters. Generally, data collection intensity peaked in the mid-1970's, in the sense of the maximum number of sampling programs being underway and the densest network of stations, and has been declining since. However, there have been notable increases in the collection frequency of hydrographic parameters and trace parameters. Figure 5-79 shows both the 1970's peak and the recent increase in salinity and DO (temperature and pH being almost identical). Figure 5-80 shows the collection intensity of the principal state monitoring programs that determine these hydrographic parameters, and clearly demonstrates that the increase since 1980 has been due to the intensified data-collection of the Texas Parks and Wildlife Coastal Fisheries program. The importance of this data collection enterprise, in terms of the raw numbers of observations made, cannot be overstated. The activity of the TWDB-sponsored Bays and Estuaries Program in the early 1970's should be especially noted. These data were collected largely by contractors to the TWDB, including U.S. Geological Survey, investigators at the Marine Science Institute, and various private consulting companies.

Conventional water contamination indicators are shown in Figs. 5-81 and 5-82. The decline of samples since 1993 is due in part to the pipeline problem of sample processing at TNRCC, which has a backlog of many months (see Section 5.3.2 below). The intensity of coliform data collection appears to have generally increased since the 1960's. This is in part due to the data collection by the state agencies, especially TDH, see Fig. 5-83, but in part to the loss of data from Nueces County Health Department taken prior to 1975. Also, fecal coliforms have replaced total coliforms, and while the coliform record encompasses nearly 40 years, it is really only half that good, since there is no proxy relation between the two coliform indicators. Nutrients are shown in Figs. 5-84 and 5-85. The peak in the 1970's is due primarily to the activities of TWDB Bays and Estuaries Program. Note the decline in the most recent five years. The most surprising aspect of the turbidity-related measurements, Fig. 5-86, is that the most important parameter TSS is measured on a long-term basis only by the TNRCC. Note also that NTU's have replaced JTU's in the historical record. (See Section 2.3 above.) The increase in the frequency of trace constituent analysis, noted above, was due to increased interest in a wide spectrum of metals and organic toxicants coupled with a quantum increase in analytical methodologies (e.g., mass spectrometry). This is exemplified by the parameters of Fig. 5-87, showing one metal, one pesticide, one PAH and total PCB's. Most of the parameters in the CCBNEP list were introduced by this same technological innovation, because the GC/MS methodologies permit a large generation of parameters from a single sample/procedure.

As salinity and temperature are the most easily measured variables, they represent the densest and longest data record. Even at this, past sampling practice does not generally permit analysis of time scales of variation shorter than a few days. Moreover, the extant measurements are nearly all for daytime hours, which must be considered a source of potential bias in the data. For salinity, in

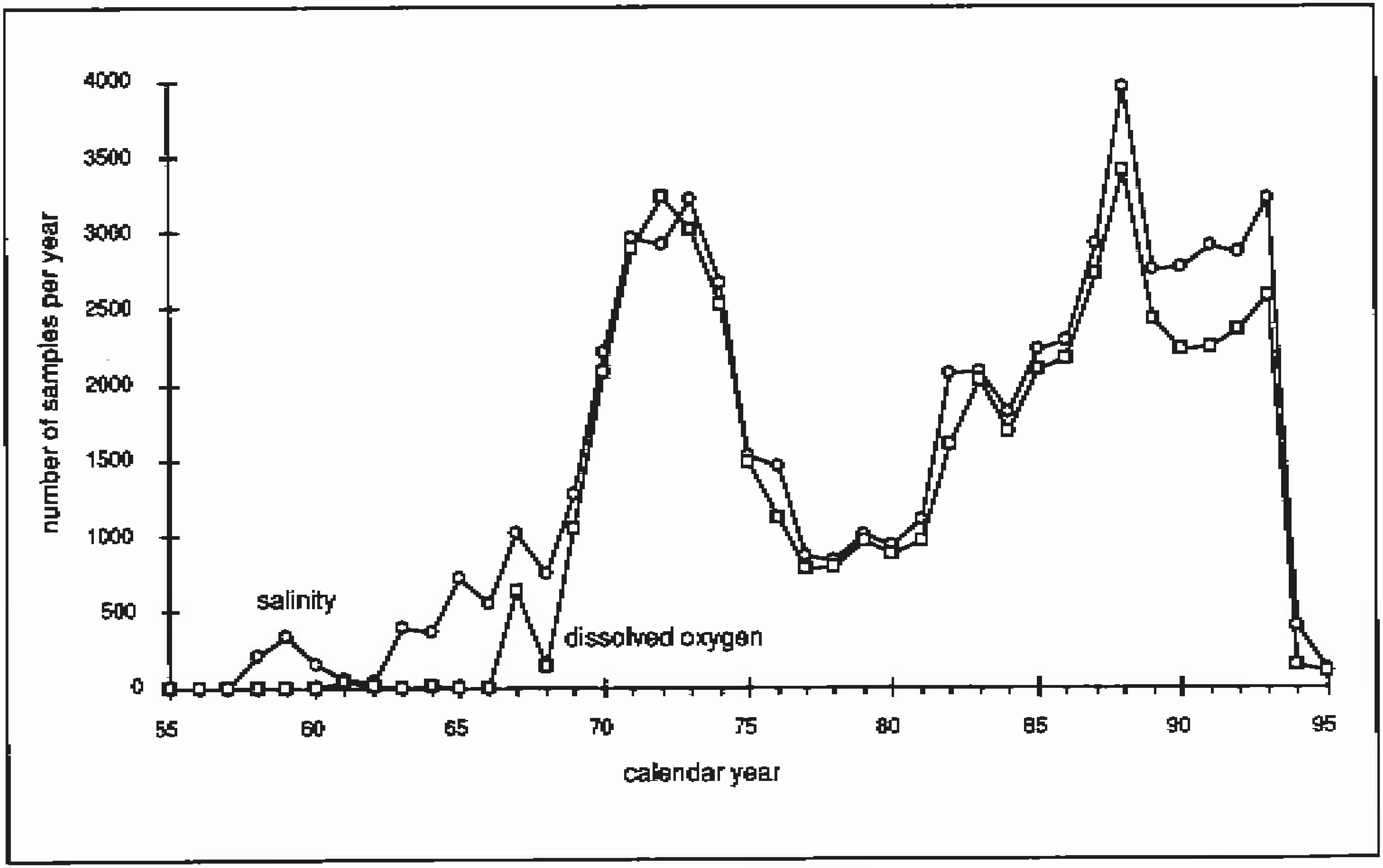


Figure 5-79. Data base annual measurements 1955-95, hydrographic parameters

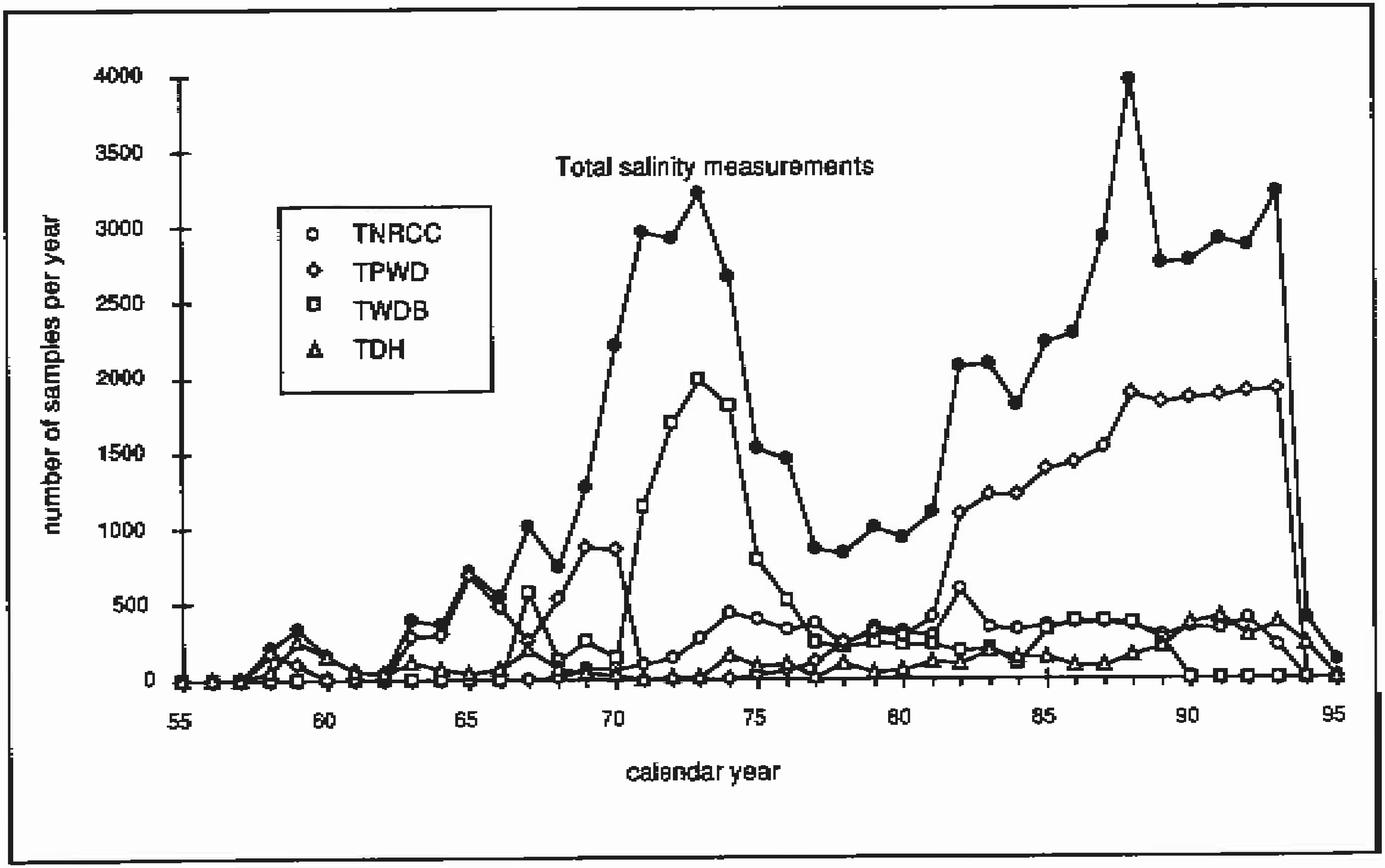


Figure 5-80. Data base annual measurements 1955-95, salinity by project

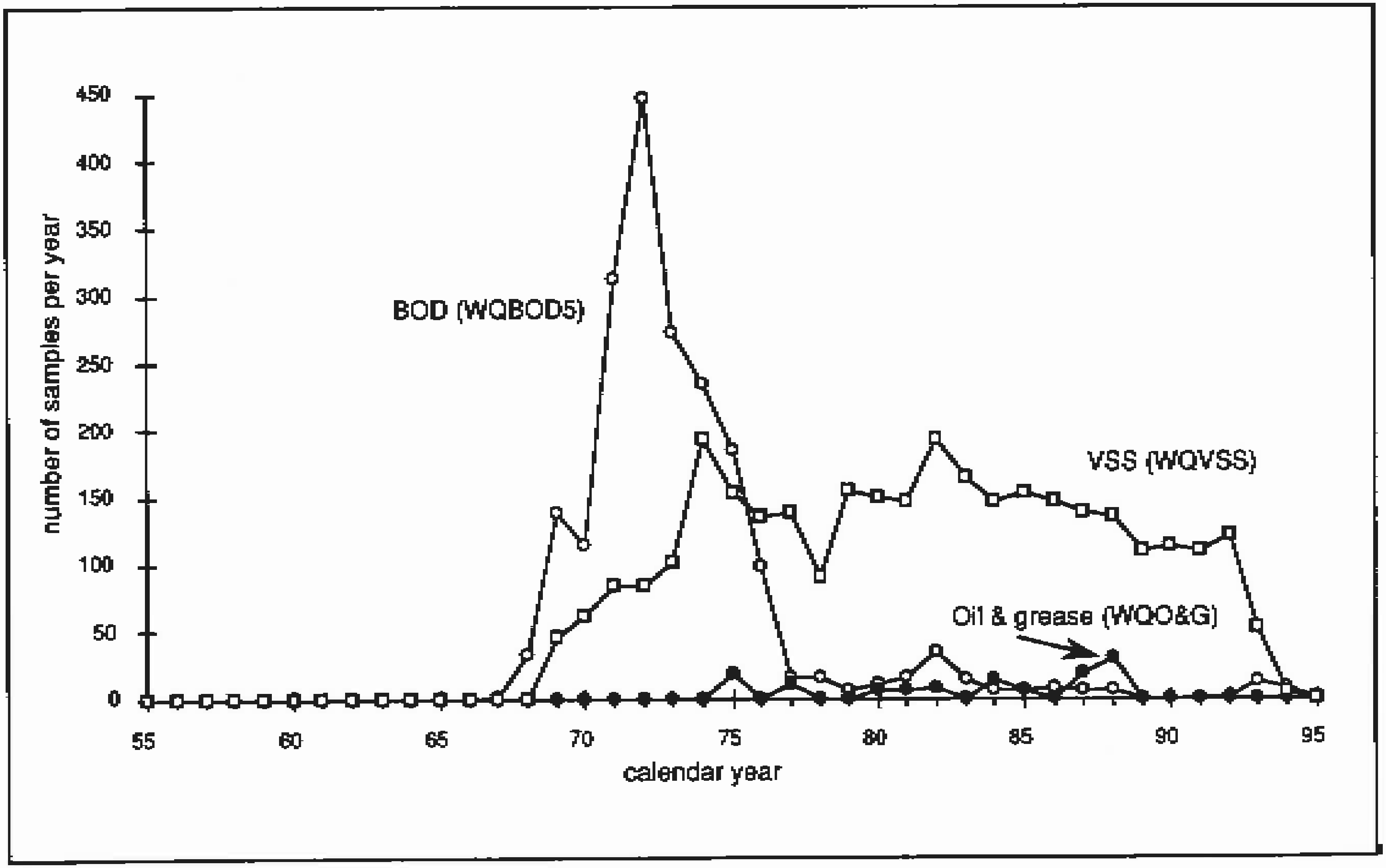


Figure 5-81. Data base annual measurements 1955-95, water-quality indicators

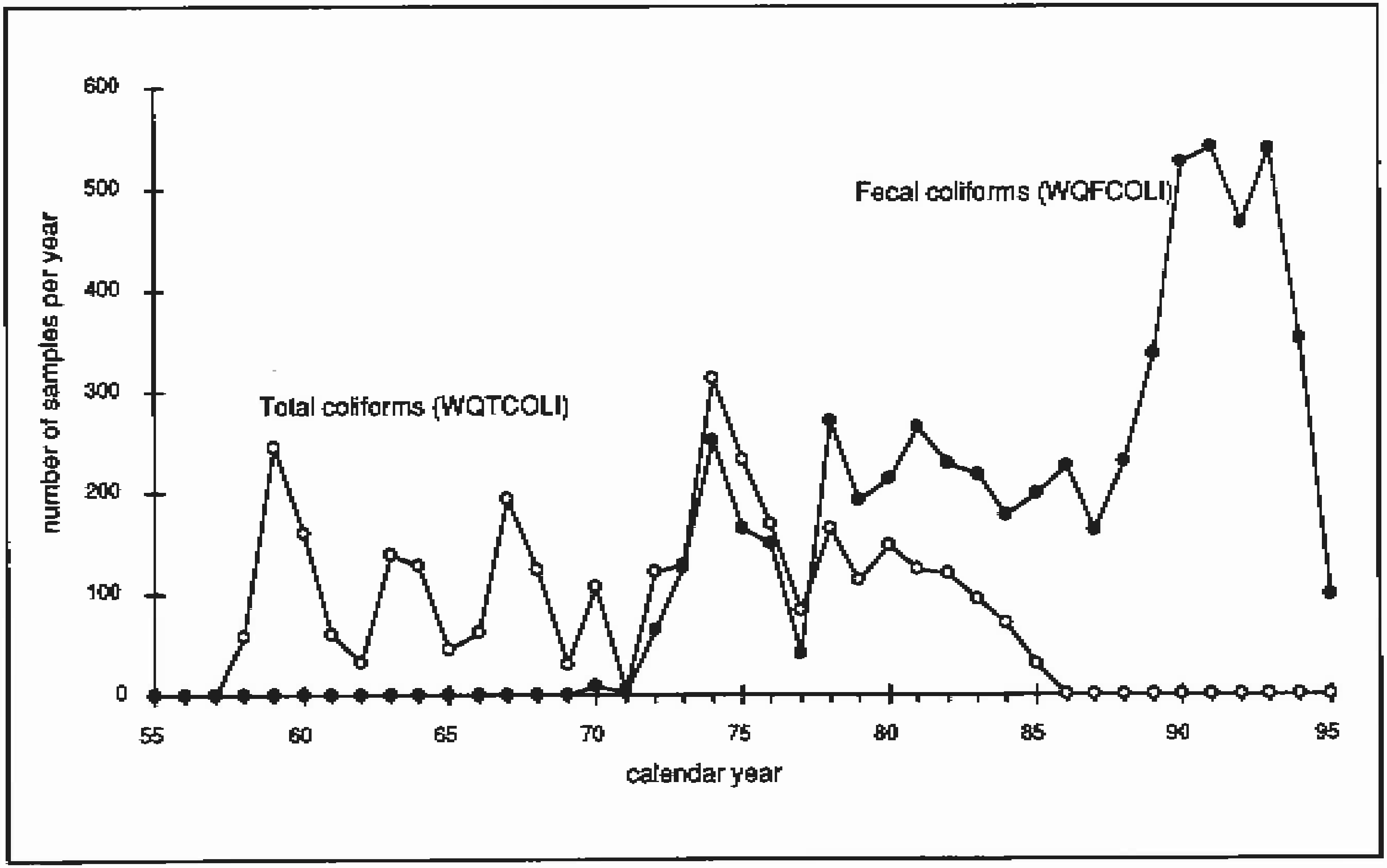


Figure 5-82. Data base annual measurements 1955-95, coliforms

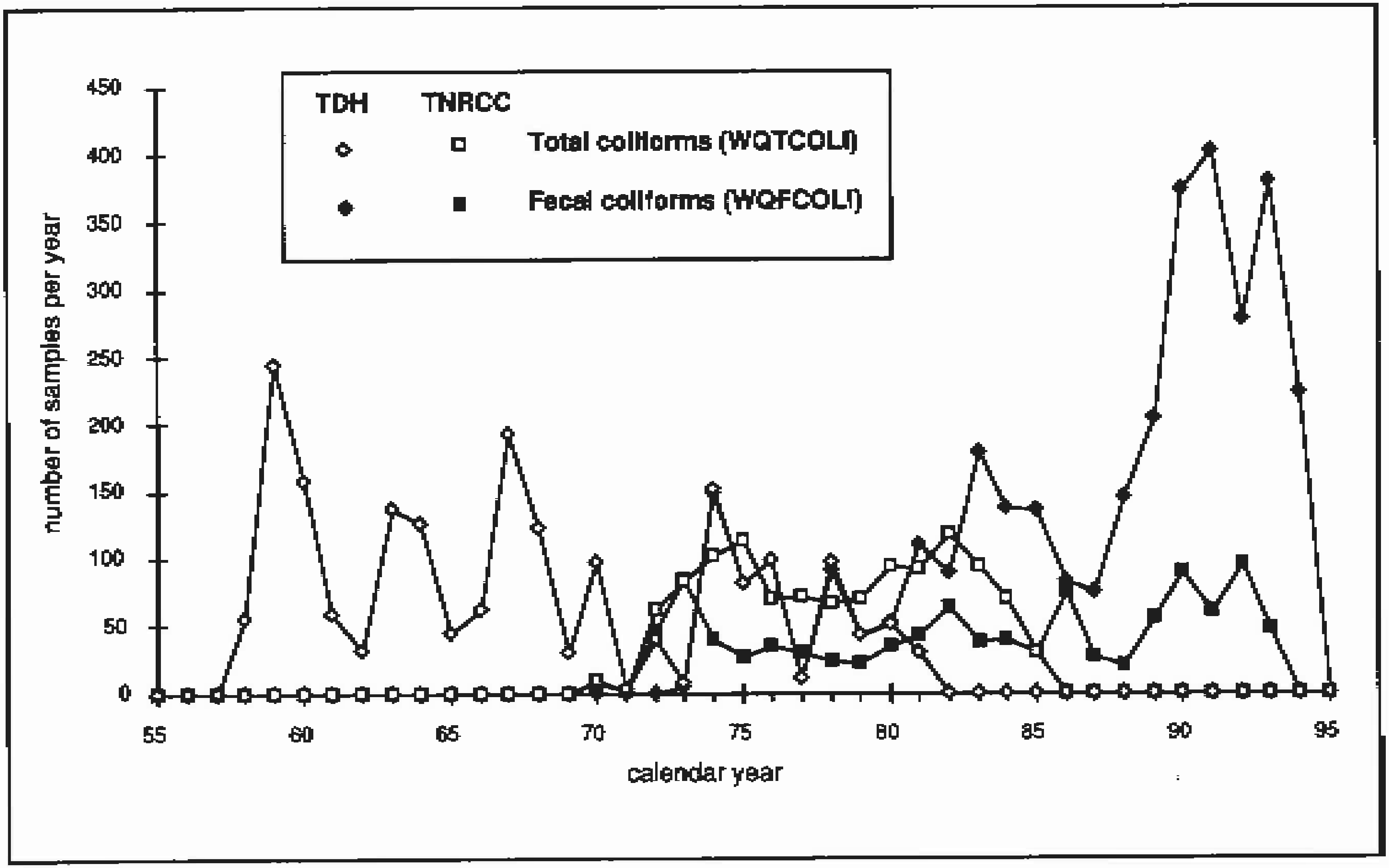


Figure 5-83. Data base annual measurements 1955-95, coliforms by state project

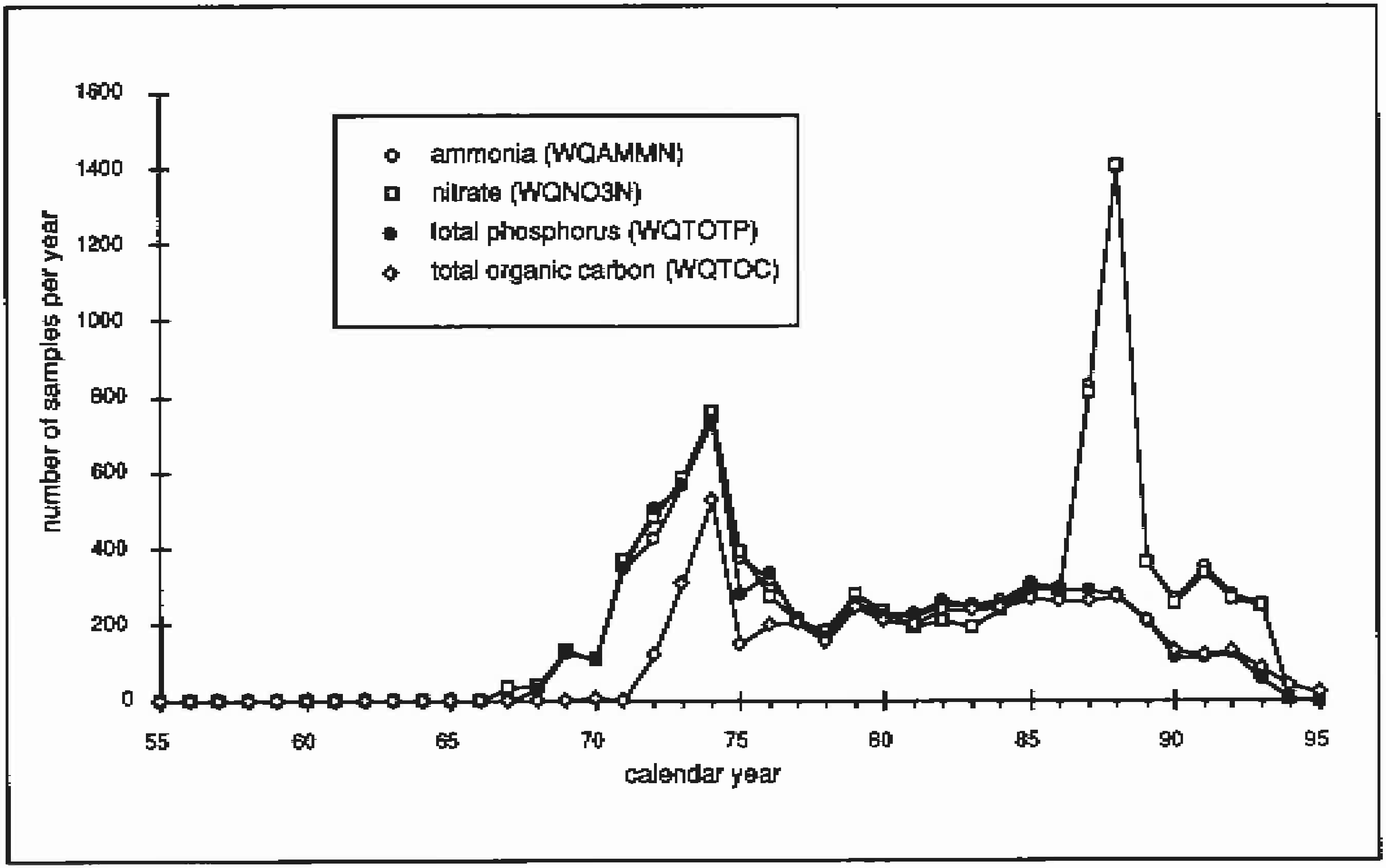


Figure 5-84. Data base annual measurements 1955-95, nutrients

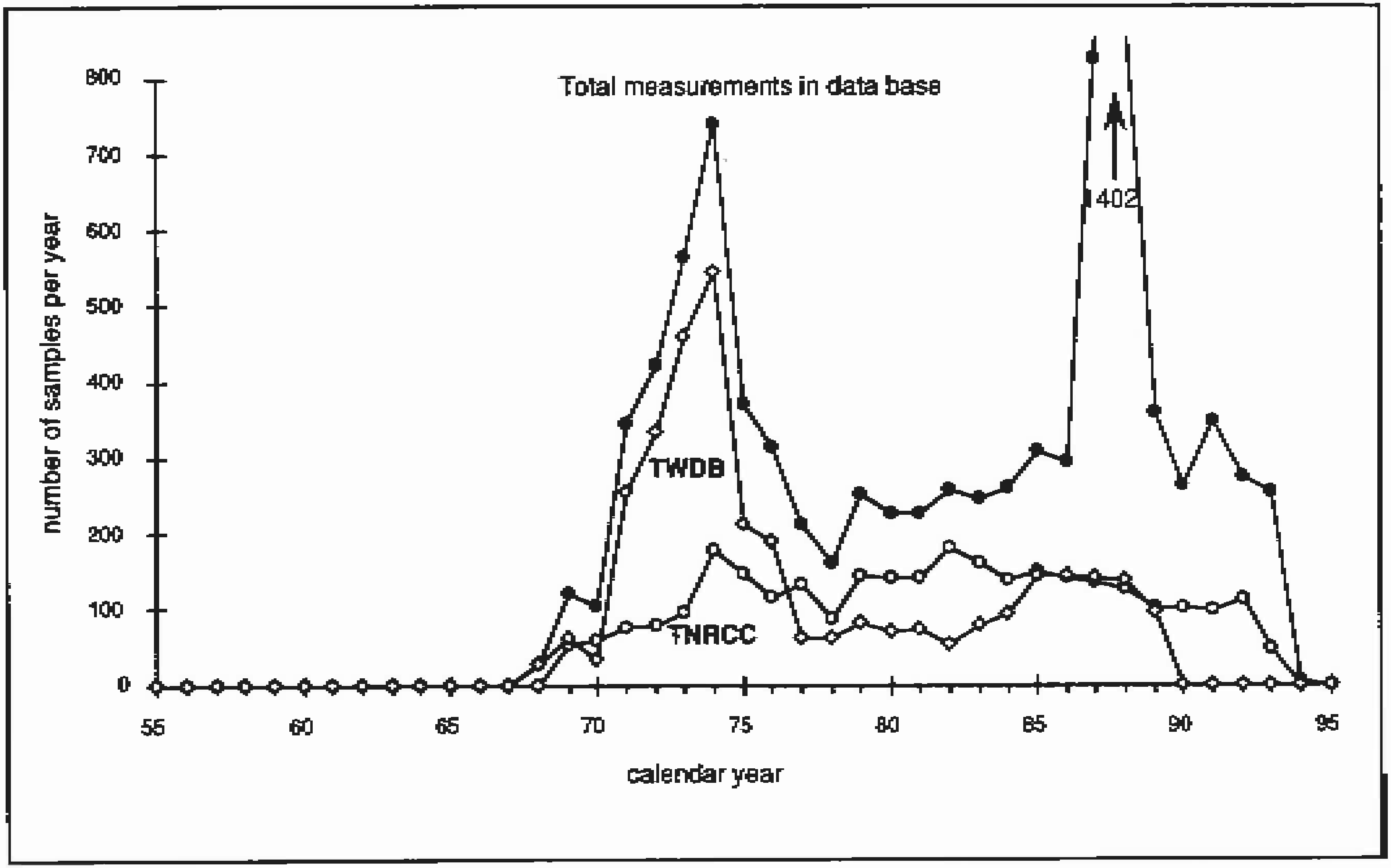


Figure 5-85. Data base annual measurements 1955-95, ammonia (WQAMMN) by state project

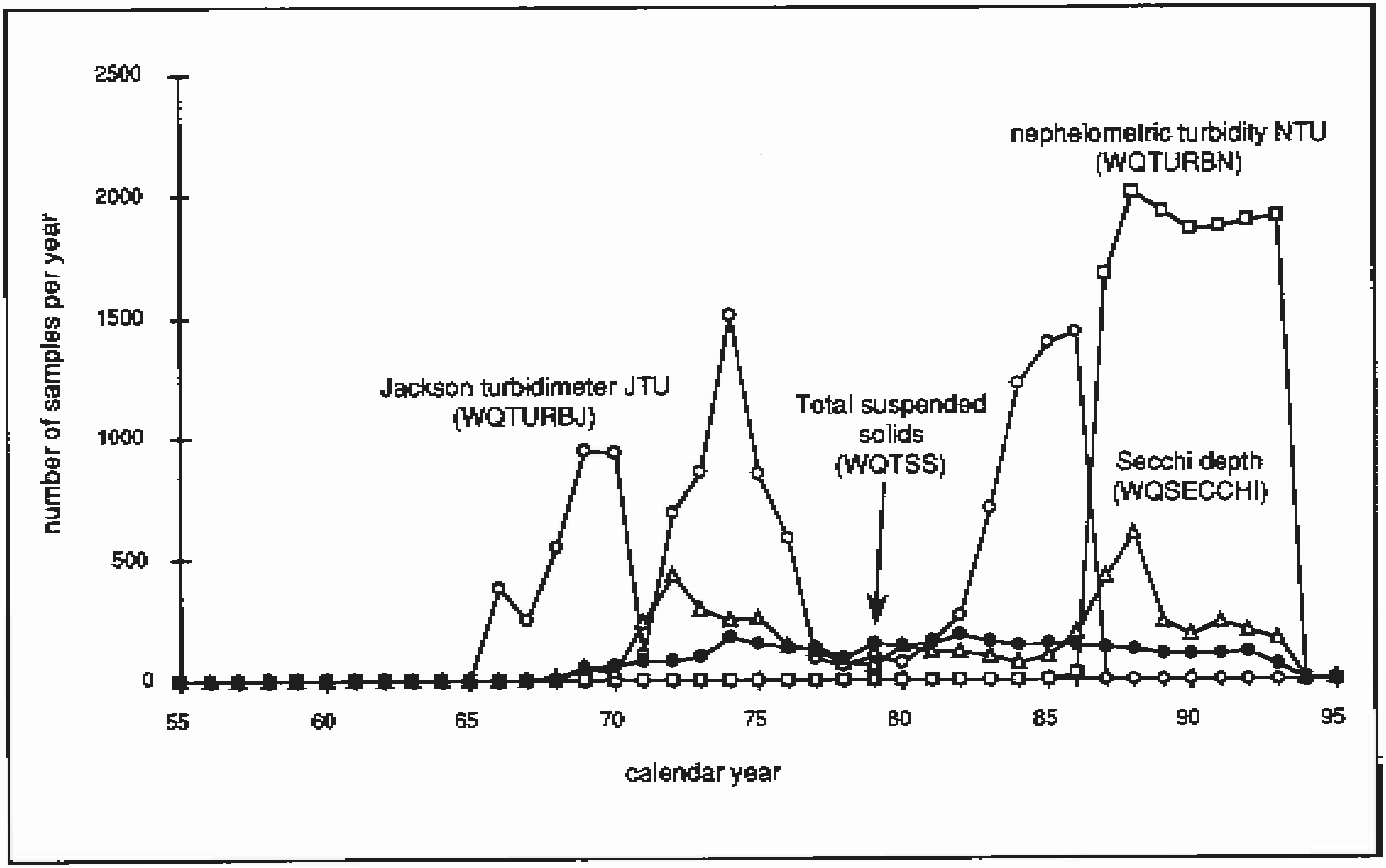


Figure 5-86. Data base annual measurements 1955-95, turbidity parameters

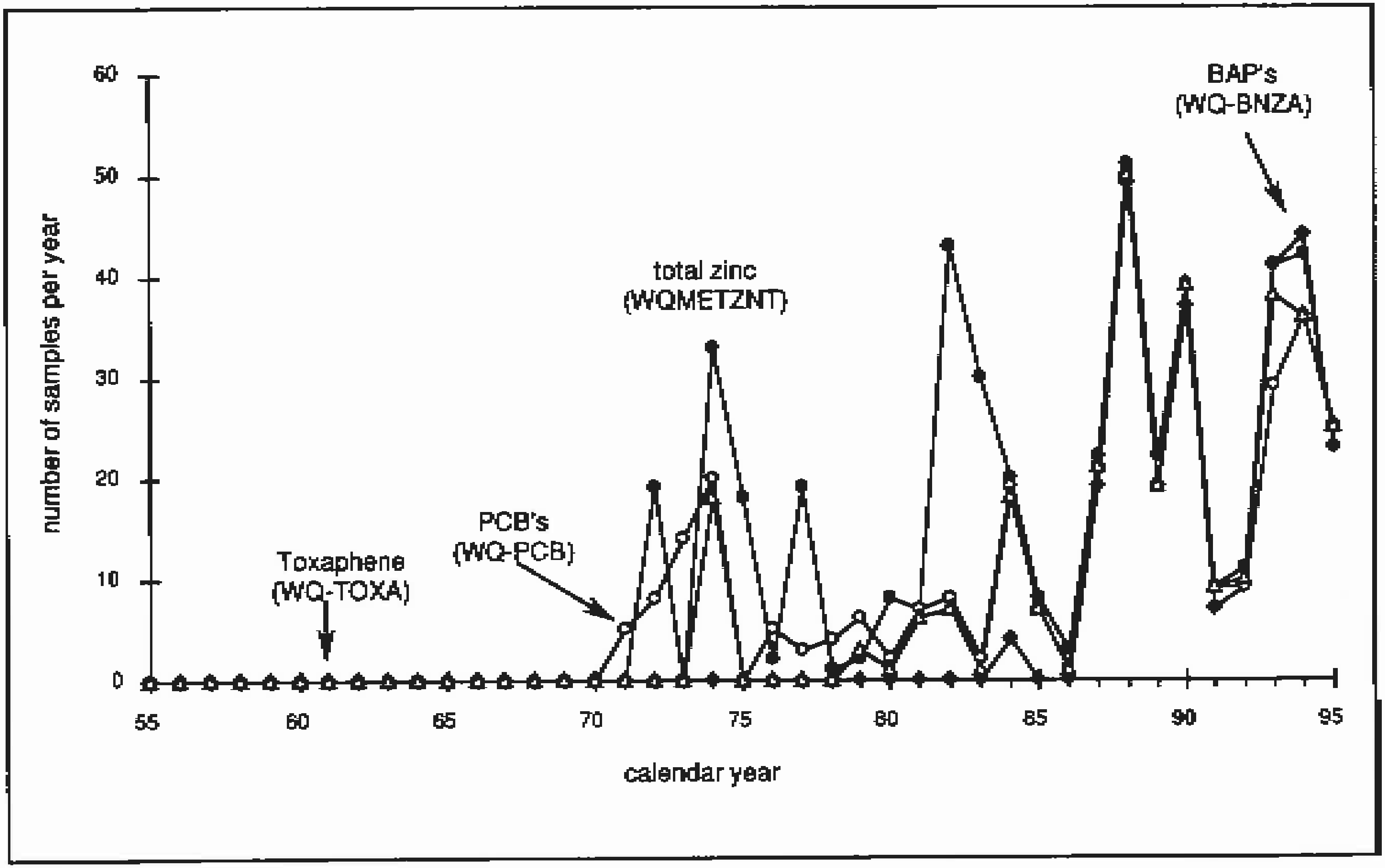


Figure 5-87. Data base annual measurements 1955-95, metals and toxics

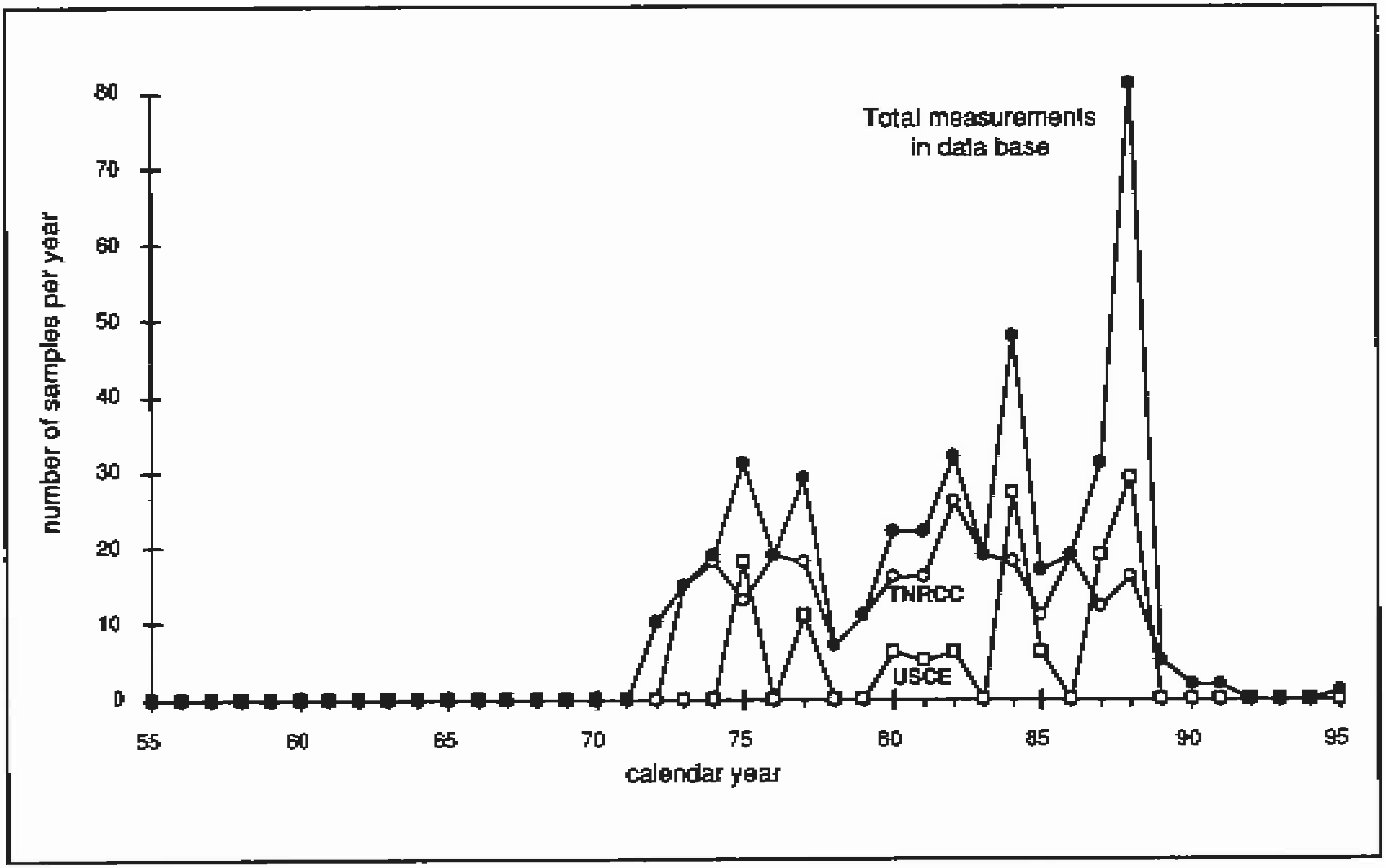


Figure 5-88. Data base annual measurements 1955-95, sediment oil & grease (SEDO&G)

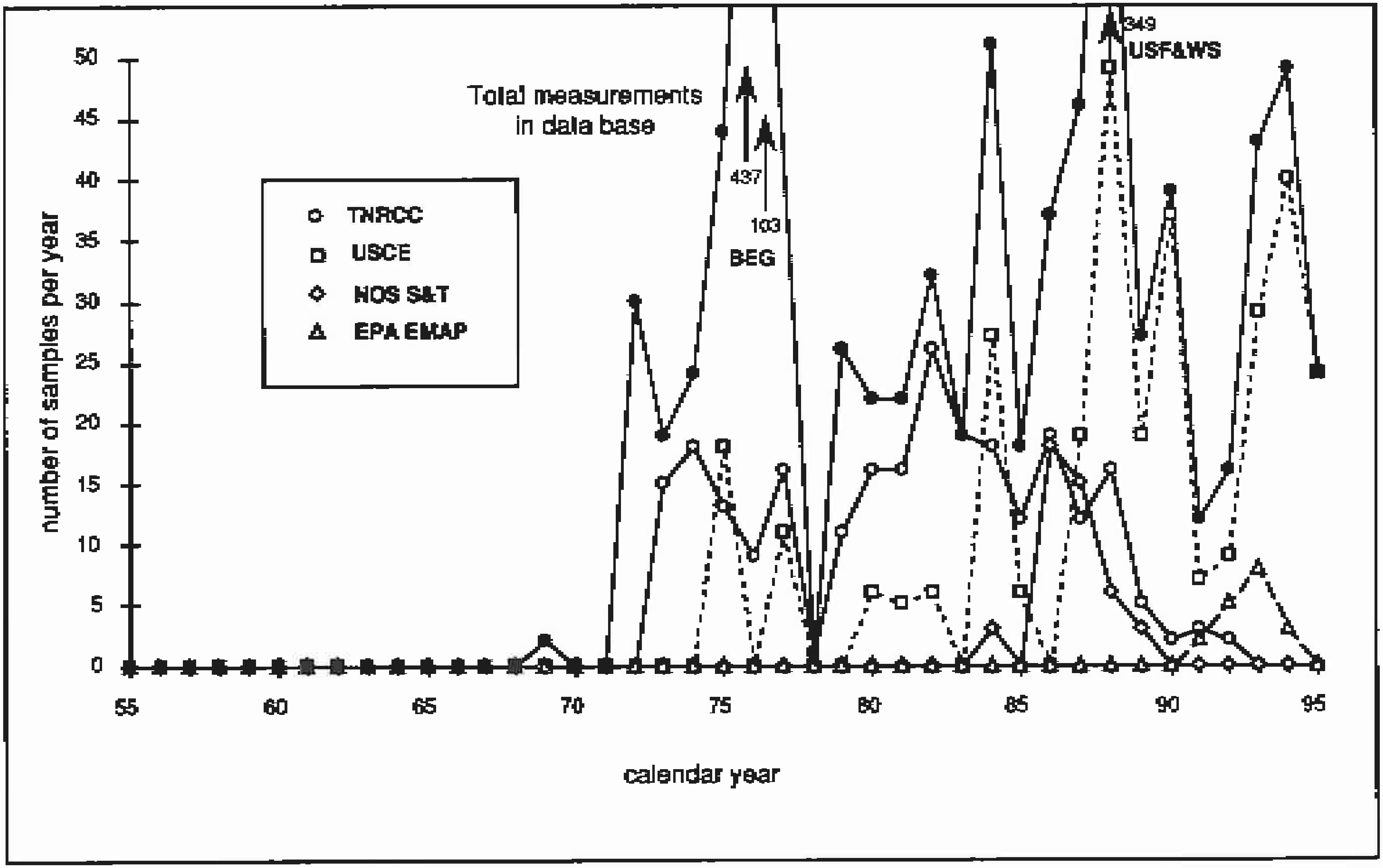


Figure 5-89. Data base annual measurements 1955-95, zinc in sediment (SEDMETZN)

some areas of the bay (depending upon the presence of steep gradients), there is a tidal oscillation that is generally unsampled, but certainly a source of variability. For temperature and dissolved oxygen, especially, there is a known diurnal variability which is virtually unsampled by routine observations in Corpus Christi Bay. The use of automatic data logging and electrometric sensing now permit the recovery of nearly continuous, fine-scale time signals of these parameters, and should be incorporated into routine monitoring of the bay. Their present use poses its own set of problems, discussed in 5.3.2 below. NB, such data acquisition does not replace routine sampling, since routine sampling provides far better spatial continuity and is not subject to vandalism and sensor degradation (see 5.3.2), which plague automatic monitors.

One of the central problems in constructing a sufficiently dense and long-term data base for analysis for the conventional parameters is the inconsistency in measurements and analytical methodologies from one program to another. Some programs emphasize COD and sulfides, say, while another examines phytoplankton and TOC, and a third may analyze BOD and chlorophyll-a. (Certainly, research demands that the utility of different indicator variables be explored, and the specific objectives of a given program may necessitate non-conventional analyses. On the other hand, the major investment is usually in acquiring the sample. Some consistency with other sampling programs could be reasonably attained at minor cost.) The PAH suite obtained by the Corps of Engineers is different from the PAH measurements of the TNRCC; the National Ocean Service Status & Trends Program obtain a different suite from either of these. And so it goes. The net effect is limited data coverage in a specific parameter that makes spatio-temporal analysis uncertain. Even for salinity, were it not for reliable proxy relationships, our ability to synthesize a comprehensive data set would be seriously truncated.

In recent years, there has been a shift of emphasis from rather gross and imprecise measurements to more precisely defined analytes. In most cases, this has involved a replacement of the old parameter with the new, so that the data record for the older parameter terminates, the data available for the new parameter is extremely limited, and there is no information as to the probable association between the two. Examples include the replacement of total coliforms with fecal coliforms, of Jackson turbidimeter measurements with those from nephelometers, and of oil & grease with total PAH's and then with specific PAH's such as napthalene, acenapthene and fluoranthene. (Further, in some programs, the specific PAH's vary from run to run.)

One of the principal properties of the water of Corpus Christi Bay is its turbidity. Suspended solids are particularly important in characterizing water quality because of the rôle particulates play in habitat quality, and in the sorption of nutrients and contaminants on the finer particulates. However, some programs do not obtain any measure of turbidity, and the only two that do obtain suspended solids (TNRCC and OxyChem) do not measure the grain-size distribution. Even a simple sequential filtration to determine partitioning of clays-and-finer would be of immense value in interpreting the data. The understanding of the behavior of most nutrients, metals, pesticides and priority pollutants is limited by the lack of information on suspended solids in the water column.

Metals data are dominated by total (unfiltered) analyses. So little measurement has been made of the dissolved phase that no characterization or trends analyses are possible. This practice is perhaps not inappropriate because of the known affinity of trace metals for particulates, but underscores the problem of not having paired measurements of suspended solids to which the total concentrations could be related. Future sampling should include routine measurement of suspended solids with every metals sample. It would be even better to include a determination of grain-size distribution, as noted above. Much of the historical data for metals has been corrupted by inattention to detection limits. Frequently detection limits are reported in error (perhaps not determined as a part of the analysis) or, worse, zero values of concentration are reported.

Pesticides and other organic contaminants are a relatively recent addition to the suite of measurements, and the water-quality data base is presently inadequate for any detailed analyses. The best record is the extended DDT, obtained by combining reported "total" values with those estimated from the pp'-DDT isomer using the proxy relation (2-10). Even at this, there are only 315 observations of which only 6 are above detection limits in the entire bay (excluding the zero values discussed above). Interpretation of organic-contaminant data generally is based upon normalization to organic carbon (e.g., Karickhoff, 1981, Moore and Ramamoorthy, 1984a). For Corpus Christi Bay, there are practically no paired measurements of organic carbon and organic contaminants.

Sediment data is extremely limited for the bay. This is unfortunate because (1) the shallow nature of the bay would suggest that sediment interactions should be a signification factor in the quality of the overlying water and its habitat value, (2) sediment is considered to be a long-time-constant integrator of bay quality, compared to the variable and evanescent nature of the overlying water. While the number of observations given in Table 5-3 might appear to be large, for the conventional and metals parameters, the densest data sets, they reduce to on-the-order-of 50 observations (per parameter) per year over the period of record (and this is misleading, since the samples concentrate in a few specific years), see Figs. 5-88 and 5-89. If distributed uniformly over the study area (which they are not), this amounts to one sample per 50 square miles per year. For many metals and most organic pollutants, the data base is even smaller, and, moreover, only about 10% of the measurements are above detection limits (again, excluding the anomalous zero values).

By far, the Corpus Christi Bay component has the richest data base. For most of the sediment parameters, over half of the available data are taken from this bay. The Aransas-Copano system accounts for about a fourth of the data, and the remaining fourth is divided about equally between the Baffin system and the Upper Laguna Madre. For the objective of quantifying anthropogenic impacts, especially of industrial or urban origin, this is probably an appropriate distribution, since it is in rough proportion to the probable degree of human influence. For the purpose of characterizing natural sediment quality as a basis for habitat assessments, the distribution is not appropriate. Neither Copano nor Baffin have been sampled anywhere fore anything in over five years.

The spatial distribution of data in these systems is dominated by two survey projects. The first was the Submerged Lands project of the Bureau of Economic Geology in the mid-1970's (Project 12 in Table 5-1). This collected sediment samples at approximately one-mile centers throughout the Gulf nearshore and estuarine coastal zone. At each of these stations TOC was determined, and at about one-third of them a suite of metals was analyzed. The second project was the contaminants study of the U.S. Fish and Wildlife Service (Project 27 in Table 5-1), which obtained metals and (at a subset of the stations) organics data at stations at one-mile centers throughout the system, except for Aransas-Copano.

The most important monitoring project that routinely collects sediment quality data throughout the system is that of the TNRCC, see Fig. 5-89. USCE collects data at irregular intervals near its navigation projects. Both the NOS Status & Trends Project (18 in Table 5-1) and the EPA EMAP/REMAP Project have limited periods of record and there is considerable doubt about their continuation.

A major deficiency of the sediment data base is that there are almost no measurements of sediment texture (i.e. grain-size distribution). Many of the parameters of concern, such as heavy metals and pesticides, have an affinity for fine-grained sediment, and moreover probably enter the system through run-off, also the source for most of the fine-grain fraction of sediment. Therefore, analysis of the variability of these quality parameters in the sediment must consider the grain-size fractions. Considering that sediment texture is an inexpensive measurement, especially compared to gas-liquid chromatography and mass spectrometry, it is inexplicable that texture data has not been routinely obtained for sediment samples. The BEG Submerged Lands Project and the EMAP programs are the only instance in which texture and metals data were both obtained. Only the BEG collected a respectable spatial density of samples, but even in this program only a minority of the samples were *paired*, i.e. chemical and textural analysis run on aliquots of the same sample.

Finally, we must note that maintenance of a monitoring project within a limited budget and resources represents a compromise between station density, temporal frequency, and the extent of the suite of analytes. Cost for all three have been increasing, the last due to more precise and expensive laboratory methodologies. Within recent years there has been a concomitant decline in all of these, especially in the spatial and temporal intensity of sampling. This is exemplified by Fig. 5-90, showing the historical discontinued stations of TNRCC and TDH, as well as the current (1994) stations.

5.3.2 Deficiencies in management of modern data

Data management is generally a shambles. Reference is made to the conclusions of Ward and Armstrong (1991) concerning data management practices and data loss in general. While these were presented with respect to the data inventory for

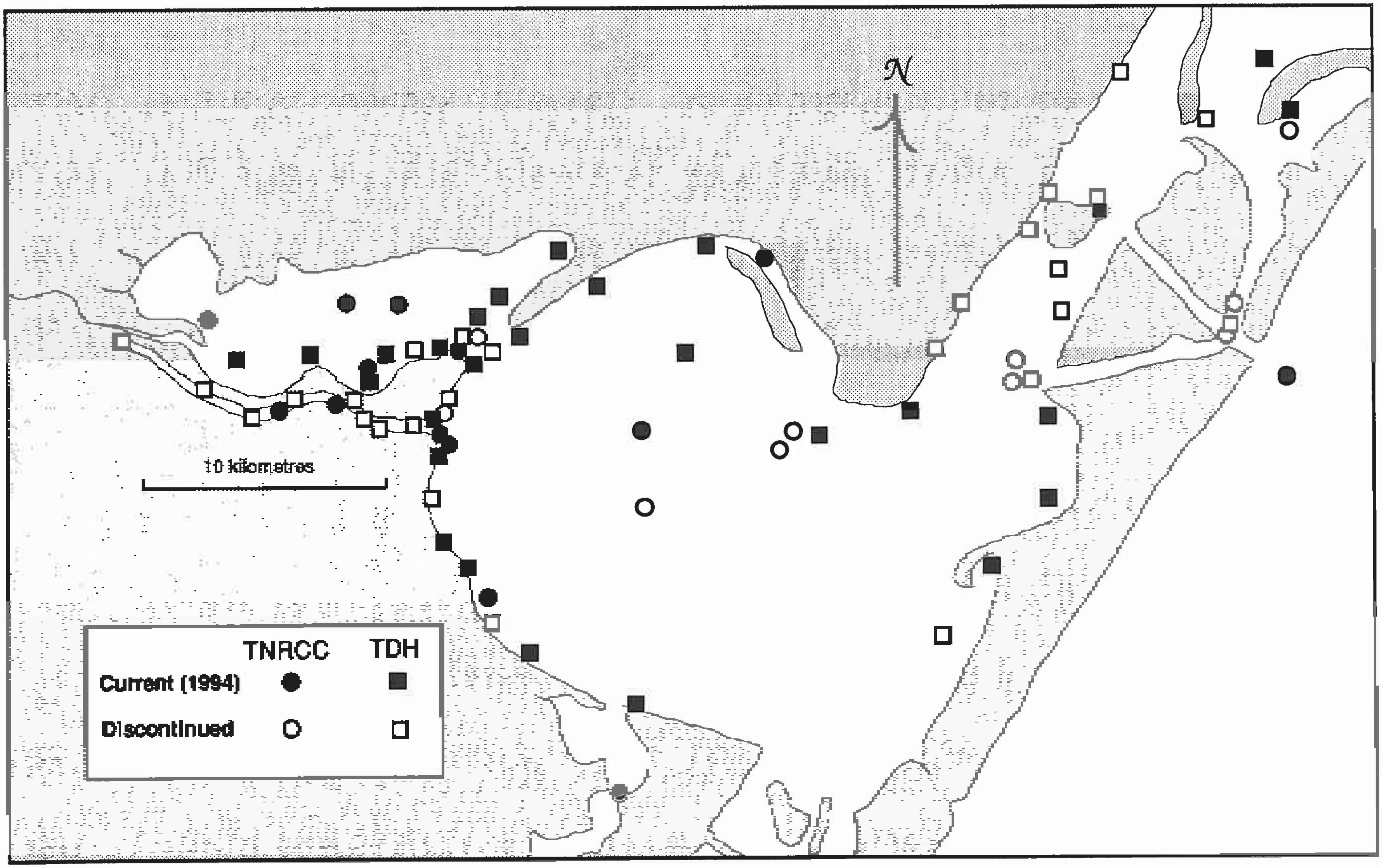


Figure 5-90. Sampling stations of TNRCC and TDH in Corpus Christi Bay

Galveston Bay, most of them apply as well to the Texas coast in general and to Corpus Christi Bay in particular. Also, while the GBNEP study was carried out over five years ago, the same problems were encountered in this study. And, as will be seen, some new ones have been added. In the present study, we identified four principal causative factors compromising the integrity of *modern* data bases. These are enumerated below with discussion and examples drawn from the present study.

1. Poor data recovery procedures, including an unwarranted trust in technology.

By "recovery" is meant all data manipulation procedures after the basic measurement has been documented by field sheet, laboratory report, or records from a data logger. These include application of necessary calibration or conversion factors, date entry (or downloading), post-processing and re-formatting. Data recovery problems arise from data-entry and data-corruption errors, and inadequate review of digitized data to detect these errors when they are capable of correction. These also include data-entry backlogs, which, though not a source of data-entry error *per se*, exacerbate the problems of detecting and correcting errors. All data recovery problems originate in neglect. From an economic viewpoint, their existence is mystifying, given the great investment of money and effort in the acquisition of the measurement itself. The necessary post-processing, checking and verification procedures to ensure that measurement has been properly entered in the data base are comparatively modest and would seem to be prudent to protect the investment already made in collecting the data.

In view of the Galveston Bay experience (Ward and Armstrong, 1992a, 1992b), it is not surprising to have identified major data management problems in the TNRCC SMN data base. We must point out that the procedures of this agency in verifying data entry and in retaining hard-copy field records, as well as the development of a new data management system, have greatly improved the situation. The problems experienced in this CCBNEP data compilation originated with the older data, for which there is probably no feasible solution at this stage. These problems include data entry errors, position errors, and incorporation of "BOGAS" measurements (see below) into the data base together with real measurements. This data program is also the prime example of data backlog. Except for the freshwater stations, virtually no data was available to this study from the SMN stations later than 1993. We were advised that this data was still in the processing "pipeline." Much of the ability to detect and correct aberrant values depends upon their timely detection. Two years is surely too long. The same backlog problem is experienced with the EPA EMAP/REMAP program in the Gulf of Mexico, for which 1994 (!) data is still being processed and Q/A'd, therefore very little could be made available to this project.

An example of the problems attending to the unwarranted trust in technology is the U.S. Fish & Wildlife Service sediment contaminants program of the late 1980's (Barrera, et al., 1995). From the standpoint of data preservation and dissemination, the USF&WS did everything right in this project–a rare accomplishment–including the inclusion of a diskette in the endcover of the report containing all of the raw data in spreadsheet format. Unfortunately there was no review of the final spreadsheet data files. We discovered that about half of the aromatic hydrocarbons from the sediment analyses and *all* of the organic analyses for the tissue samples were absent from these files. It is noteworthy that the missing data records begin at pagebreak positions in the spreadsheet. Apparently, due to a bug in the software or an incorrect print-range specification, these data were simply not transferred to diskette. The computer used for the master data file has since been purged and there were no hard copies or digital backups retained. USF&WS advises that the only way to recover the lost data would be to go back to the raw lab sheets and completely re-keyboard them. Unfortunately, this is not a lone example, but a constant frustration in the preservation of digital data bases, see Section 5.3.3.

Another instance of data-recovery problems involves automatic data-collection equipment. Experiments have been underway since the mid-1980's in the collection of time-intense water-quality data by automatic recorders equipped with electrometric sensors. The Bays and Estuaries Program of TWDB was the first agency in the Corpus Christi Bay study area to employ such robot observation systems, referred to generically as "sondes." The Marine Science Institute of the University of Texas and the Conrad Blucher Institute of Texas A&M–Corpus Christi have also experimented with robots, CBI incorporating sondes into its program of tide gauging (the TCOON system). These data sets are valuable in providing insight into a element of variation of Corpus Christi Bay that is virtually unsampled, and the surface has only been scratched in the analysis of the data. However, no sonde data have been included in the present Status & Trends data compilation. The reason is that the data is either uncorrected or untrustworthy.

The electrometric probes are prone to fouling and degradation, especially in the saline environment. Even salinity (i.e. conductivity) and temperature, the simplest of the sensors, can exhibit substantial drift in time due to these effects. Dissolved oxygen is even less reliable. Some sondes are self-contained with a built-in data logger. Sampling is activated by an internal timer, whose accuracy can degrade over time. If care is not taken in operation of the sonde, the time part of the data measurement can drift, accumulating an error of as much as several days after a one-month service period. (Many of the CBI sondes, e.g. those of CBI, are remotely interrogated and the data is directly telemetered to a permanent facility from the sonde, so time drift is eliminated.) The keys to developing a reliable record of data from a robot sonde are frequent maintenance, careful pre-deployment and post-deployment calibration, and diligent data scrubbing (i.e., review of the sonde records and detection of aberrancies).

Despite the effort that has been invested in capital equipment, mooring and retrieval, and downloading procedures, none of the data records available to this project could be considered reliable. In some cases, the information potentially exists to validate the sonde data, but the data logs have not been corrected. The TWDB performs measurements with a separate instrument when the sonde is deployed and when it is retrieved. These measurements can be used to "calibrate" the raw data, thus correcting for sensor drift. (This assumes the drift rate is linear, but that is another matter.) Only 14% of the TWDB sonde data files have been corrected in this manner. One might reason that since the sondes are checked and calibrated before deployment, at least the first few days of the record could be expected to be accurate. However, this calibration apparently does not include adjustment of the instrument to agree with the standard, and comparison of the TWDB-corrected records versus the raw data quickly dispels this hope, see, e.g. Fig. 5-91. Moreover, an examination of the "corrected" files that are available shows that other sorts of problems, such as aberrant values, have not been removed. Fig. 5-92 shows one "corrected" record from Nueces Bay with obviously incorrect values. The CBI data does not appear to be in any better shape. While calibration data are noted by CBI staff, none of the records have been corrected; also, what may be major calibration problems have been discovered recently and are presently being investigated (N. Krause, CBI, pers. comm., 1996). A comparison of the raw records for TWDB and CBI sondes located in proximity at the JFK Causeway is shown in Fig. 5-93. Not only do the raw values differ substantially (which is to be expected), but the time responses of the two are uncomfortably dissimilar.

2. Failure to differentiate between the archival functions and the analytical functions of a data base.

In Section 2.1 was displayed the presence of BOGAS data in the older salinity data from TNRCC SMN. The reader may infer that the laboratory involved in those years, Texas Department of Health, has now been exposed in some sort of fraudulent practice. This would be an unfair and inappropriate judgment. Rather this is a prime example of a conflict arising from forcing a data base to serve both an *archival* and an *analytical* function (see Section 5.2 above). On the one hand, the SMN is a permanent digital archive for all water-quality measurements performed by the TNRCC and predecessor agencies. On the other hand, this data base is the foundation for various analyses of water quality, including statistics and model validation, carried out by the same agency. To satisfy the archival objective, the actual measurements must be preserved, without any modification, even conversion of units (which can distort the precision of the original measurement), certainly without the supply of BOGAS data. To satisfy the analytical objective, continuity in the suite of measurements in both space and time is necessary to maximize the available data base, which requires consistency in variables reported and their units. So long as the same variables are measured in the same units, there is no conflict between these objectives. Once the suite and/or units are altered, then the conflict arises.

Over the past three decades, there have been many modifications to both the suite and units in water-quality surveys, and the TNRCC in trying to have its SMN data base satisfy both objectives has compromised its archival integrity. In the early period of data collection, the lab conductivity was almost always available, so it is easy to see that it would be desirable to maintain a continuity of record by supplying a "lab conductivity" when the actual measurement was chlorides or field conductivity. In all likelihood this would have been done by hand calculation either in the lab report or at the data-entry stage. The problem with the SMN practice is not this entry of BOGAS data *per se*, but failures to flag BOGAS data and to check the calculation. While this practice is no longer observed in the

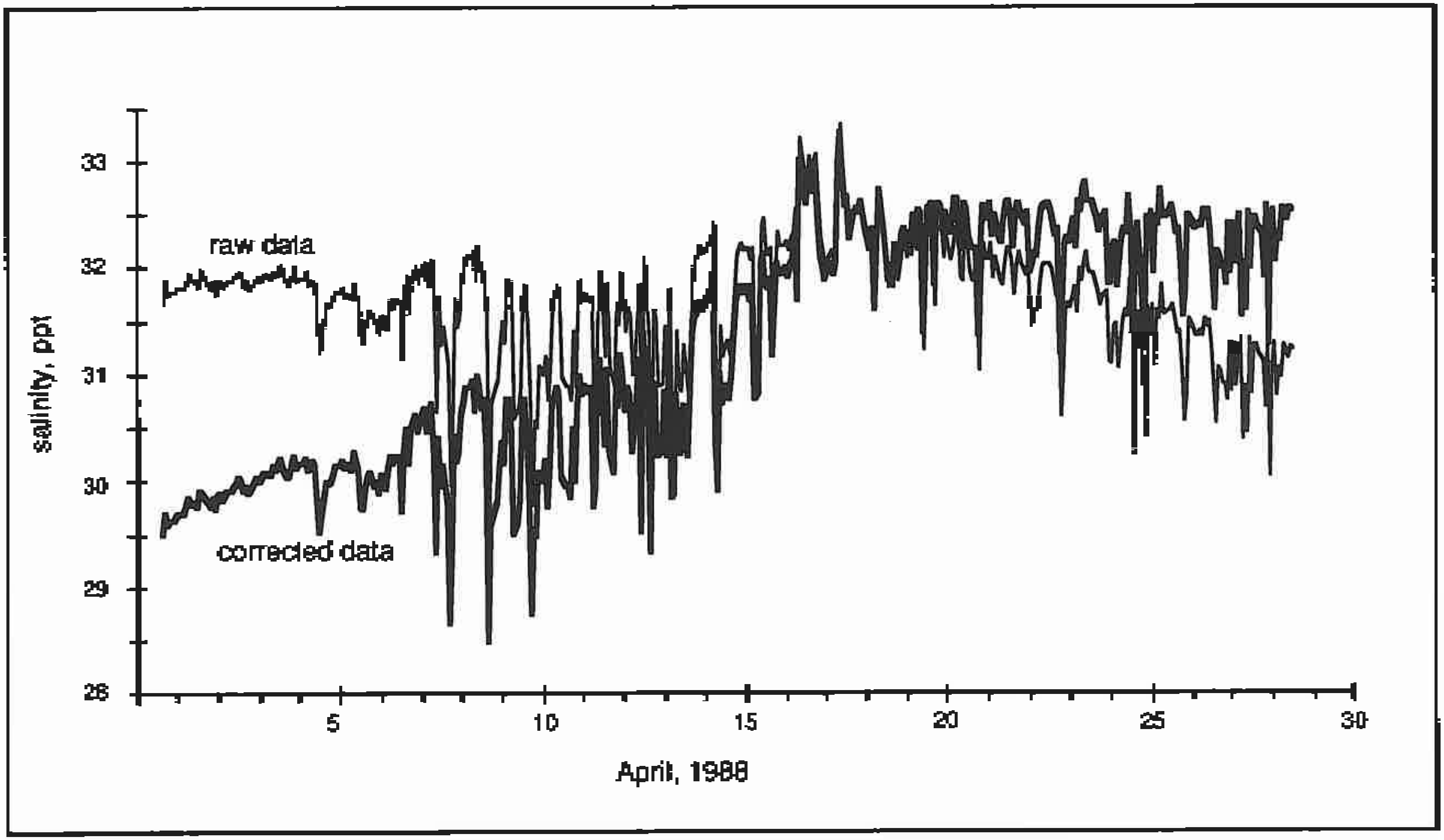


Figure 5-91. Salinity record from hydrosonde in Corpus Christi Bay near Ingleside, raw data and corrected

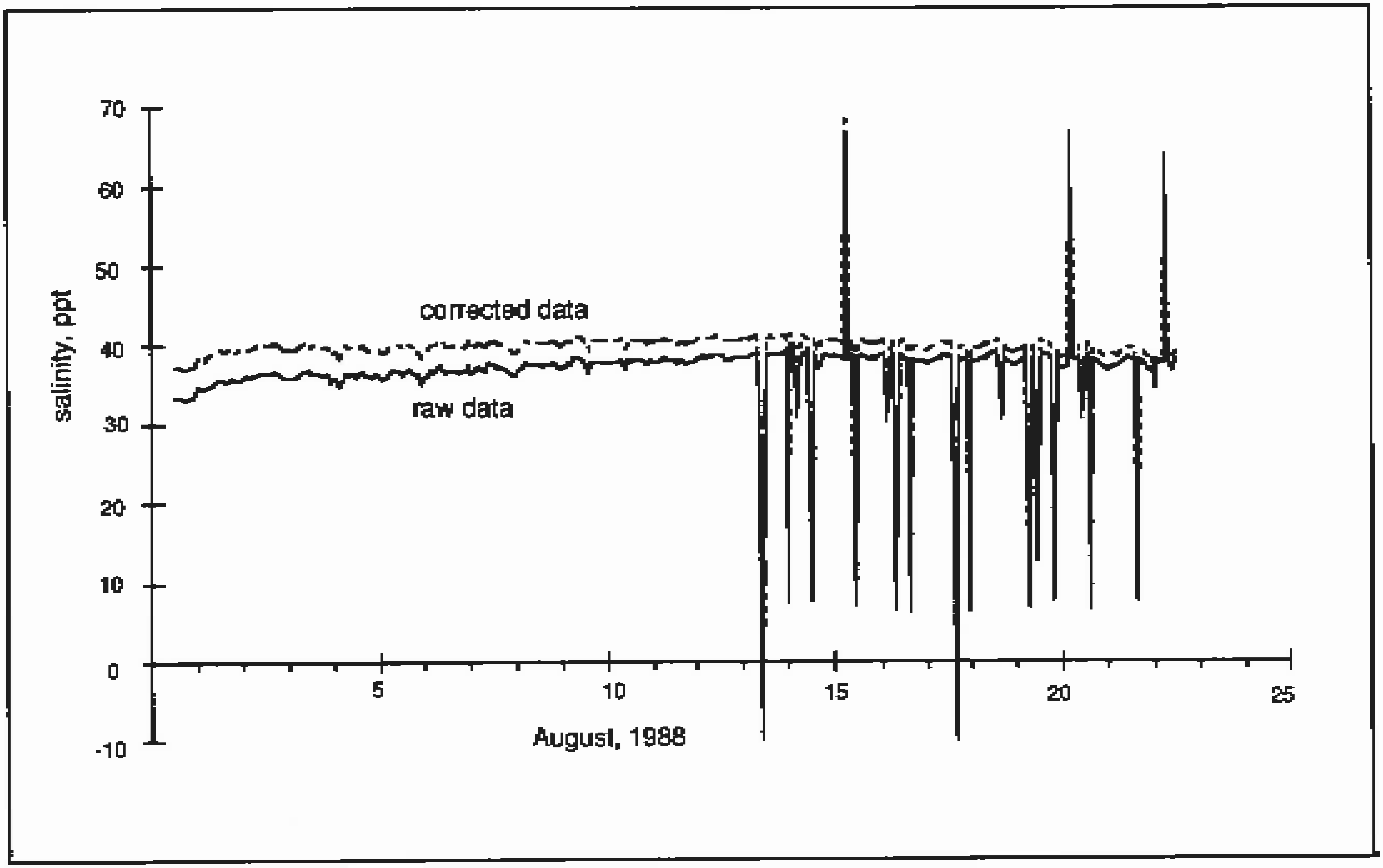


Figure 5-92. Salinity record from hydrosonde in Nueces Bay, raw data and corrected

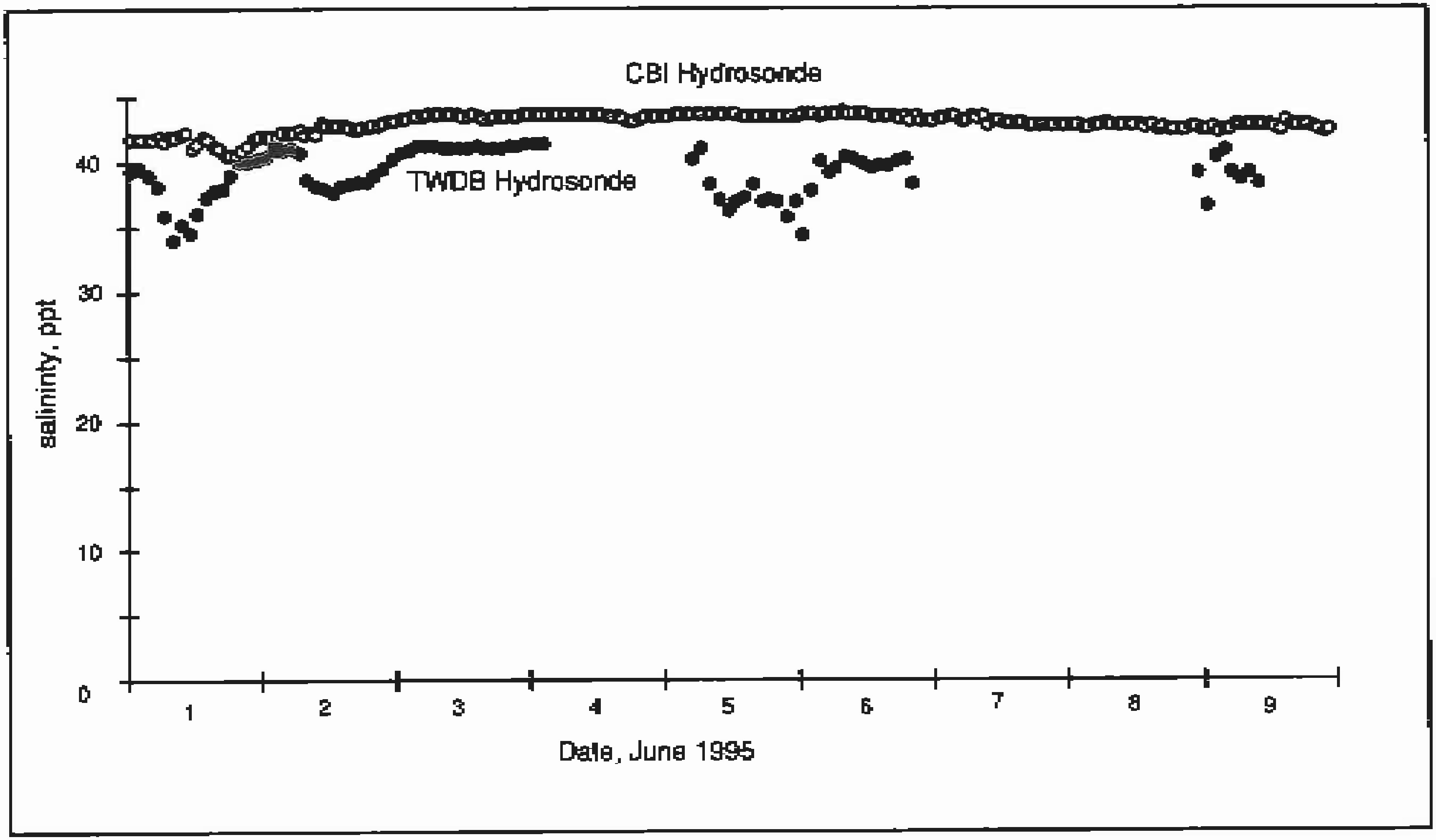


Figure 5-93. Salinity records from hydrosondes moored in GIWW at JFK Causeway

present TNRCC procedures, it is instructive to note them here to document some of the screening and data-rejection decisions that were made in the analyses for this project, and to caution against such practices in other endeavors.

Another example of the same problem is the frequent occurrence of zero concentrations for organic analytes in both the TNRCC SMN and the TWDB Coastal Data System, probably due to the replacement of censored (BDL) data by a zero value, which is indeed one of the most common means of treating censored data of this sort for *analytical* purposes. While this practice has stopped since the mid-1970's, it nevertheless illustrates the practice of altering the archived measurements to achieve an analytical objective. Still another example is the incorporation of data from other agencies into the TWDB CDS to support the *analytical* function of that data base. Each of these measurements is assigned a station location using the TWDB line-site system that is nearest the originating agency's station location. Once this "re-located" non-TWDB measurement is added to the system, there is no means for identifying it as having originated from another agency. We suspect that some of the data in the CCBNEP compilation may in fact be duplicated measurements, from both the TWDB and another agency, but because their geographical positions differ by as much as a mile, we cannot be justified in removing them.

3. Lack of a suitable archival structure for a data base.

This includes *ad hoc* filing procedures, changes in software basis without updating, and failure to provide for data preservation. Two examples from the Corpus Christi Bay compilation will be given (though there are others). The Corps of Engineers presently maintains its O&M data in three different formats. The older data are in hard-copy format only. Data from the late 1980's are in LOTUS spreadsheets (but in many cases only hard-copy printouts have survived). More recent data are maintained in yet a different data base manager. Much effort was required in this project to re-format and reconcile the data in these different formats. The second example is the Coastal Data System of TWDB. This was originally housed on the TNRCC mainframe, but with the impending demise of that system, the CDS has degenerated to a lengthy catalog of files maintained on various platforms, the files corresponding to older surveys of the Bays & Estuaries Program, to data from past contractors, and to data from cooperative studies of several state agencies. Moreover, no single individual seems to have control or management authority. We had to make repeated, increasingly specific requests over a sustained period to obtain all holdings for the Corpus Christi Bay study area, and discovered, too late, that these did not include the intensive inflow studies because these were in the possession and use of yet other members of the TWDB staff.

This is a common problem for those agencies for whom data collection is not the primary responsibility of the agency, but rather is a supporting function. This unfortunately describes all academic research, all local and regional agencies, and most state and federal agencies. To design and implement a digital archival system for data requires not only foresight, but an initial investment in effort that is usually viewed as unwarranted by the agency. Only after various data files are accumulated in differing formats is the need for a basic archival structure perceived, and then the effort of retrofitting the accumulated data would represent a diversion from other, more important responsibilities. The situation is therefore self-perpetuating. It has been exacerbated by the recent proliferation of software products, and the promotional device of American software purveyors of issuing updates and new releases at a diarrhoetic rate, which has created widespread incompatibility and non-transportability of data files.

4. Poor data dissemination protocols.

In many respects this is a corollary to the previous problem of an inadequate archival data-base structure, since one can argue that a suitable data dissemination procedure is one of the requirements of such an archive. We list this as a separate problem, however, because (1) there are archival systems that are satisfactory in all other respects except they lack a means for a non-agency user to receive and/or easily manipulate a data file; (2) because most agency use of the data is internal, the lack of a suitable dissemination protocol may go unnoticed. One example from this study is the EPA-sponsored EMAP/REMAP program. Data dissemination is performed on a case-by-case basis, in which the program data manager in Gulf Breeze routes the request to the contractor in Lafayette, who performs an *ad hoc* download to diskettes for transmittal. The formats are inefficient (e.g., station location, sampling date/time, and sample depth must be searched from separate files) and vary depending upon the calendar period for the data requested.

The National Status & Trends Project of the NOAA National Ocean Service has sought to facilitate user access to the data by having a download site on the Internet. We discovered that the only way to be sure of trapping all of the data from the project area would be to download the entirety of the Gulf of Mexico holdings. The structure of this massive data file could be charitably described as quasi-random, and required manual searching and copying to synthesize the necessary information. Then the detection limits have to be separately downloaded for each data record depending upon the code identifying the laboratory. The entire process required nearly two weeks of intensive data manipulation and reformatting before we had structured files that could be worked with.

The current fad of Internet operations is beginning to improve access to large-scale data holdings. This is welcome technology. On the federal level, much of the data holdings of USGS, and some of NOAA are now available for download via the Internet. TWDB has recently began making its data files available for direct ftp access. Formatting can still be a problem, as exemplified by the NOS S&T example above. However, one must contrast this to the situation with this same NOS data system in the Galveston Bay Status & Trends Project, five years ago. The GBNEP PI's were advised then by NOS that, although the data did reside in a digital form on the NOS mainframe, it would be easier far all concerned if GBNEP keyboarded the data from the published data reports, because weeks of special-purpose programming would be needed to produce a digital copy of the Galveston Bay data set (Ward and Armstrong, 1992b). Progress is being made.

5.3.3 Deficiencies in management of historical data

Ward and Armstrong (1991) quantified data accessibility in Galveston Bay for salinity/temperature, chemistry and biological data, as a function of the age of the data, disclosing an appalling rate of data "inaccessibility" (which includes both data that is lost and whose use is prohibited) that approaches 100% for data older than the 1960's, with for practical purposes almost everything prior to 1950 being lost. Our intuition is that a similar situation obtains for Corpus Christi Bay. While in the GBNEP Data Inventory work, six major historical data sets were rescued from "the edge of the abyss" and digitized, in the present project, three such major projects were rescued:

- the TAMU program of the 1950's sponsored by Reynolds Metals
- the SWRI monitoring program of the early 1970's
- the USF&WS-sponsored study of the system in the late 1970's

We were too late for others. Ward and Armstrong (1991) identified seven principal factors that contributed to this data loss. These were found to be operating in the Corpus Christi setting as well, and are repeated here, with appropriate examples from the present study:

1. Low priority assigned to archiving and preservation of older data.

This is a reflection of human psychology. Once a project or survey is completed, there is a tendency to stack the results out of the way and move on to the next challenge. Many agencies operate under a pressure of time, which conspires against good archival practices. Some agencies have some form of data management currently in place. While this is encouraging, it is also precarious, in that these programs are sensitive to shifts in organizational emphasis. An office purge is forever.

2. Mission-specific agency operation: perception of old data as "obsolete" and archiving as an unwarranted expense.

The Corps collects hydrographic or water quality data to support, e.g., navigation projects in place or in planning. Once a condition survey has been used to determine the need for dredging, once a decision on spoil disposal is made, once a project design is completed, the data sets employed in those activities are no longer needed. The mission of Texas Parks and Wildlife is to monitor the state of the coastal fisheries. The present condition is always primary. The Texas Natural Resource Conservation Commission and EPA are concerned with the present loadings of contaminants and the enforcement of water quality standards. The level of loadings a decade ago, or even last year, are rarely pertinent to that mission. And so it goes. The value of data diminishes quickly with age in these kinds of problem-specific operations. Yet it is these agencies that are largely responsible for the bulk of data collection within the Corpus Christi Bay system.

Among historical data collection programs in Corpus Christi Bay whose data are now lost include:

- City-County Health Department (Corpus Christi-Nueces County Department of Public Health): salinity, temperature, DO, pH, coliforms, monthly 1961-1969

- City-County Health Department: miscellaneous hydrographic and chemistry data, 1940?-1975

- Pittsburg Plate Glass: Inner Harbor stations, salinity, temperature, DO, pH, monthly 1944-47

- Corps of Engineers: Inner Harbor stations, hydrography (specific parameters unknown), monthly 1963

- Corpus Christi Sewer Department, parameters unknown, 1960?-1965

- Texas Game, Fish & Oyster Commission surveys, 1950?-1970

The loss of the early City-County Health Department Surveys is particularly frustrating, as these would have extended the record back in time substantially. It is noteworthy to observe that more recent coliform data taken by CC-NCHD have also been discarded, including those of 1976, 1977, 1983, and 1984.

Loss of the above data sets could be described as resulting from "benign neglect." In recent years, a much more malevolent entity has entered the scene, in the person of the new administrative function of Records Management, especially in fashion in the state agency headquarters. These "records managers," who are typically nontechnical and have little knowledge or appreciation of the functions of their agency, adopt a ruthless philosophy toward older information: seek out all older files, and summarily destroy those for which there is no apparent use. Massive paper purges have already been rendered at the TNRCC, among other state agencies. Holdings of district offices and storage in state warehouses have been targeted. While knowledgeable local staff have resisted thus far, it is probably only a matter of time.

3. Personnel turnover, combined with little or no documentation.

Only a handful of people in an agency generally has immediate familiarity with a data base. If the data base is not currently in use, this number will decline due to turnovers. When the last of these leave, the institutional memory goes with them. This problem is most acutely manifested in the case of a single principal investigator at a university. Most of the rare data sets we succeeded in locating for this project resulted from contacting (finally) the one or two persons remaining in the agency (or in the area) that knew something about the data.

The most extreme instance of this is when the entire data-collection entity is terminated. An example is the operation of the Southwest Research Institute Ocean Science and Engineering Laboratory in the early 1970's. This agency acquired a large amount of field data during its existence, as well as obtaining copies–perhaps originals–of data from other entities in the Corpus Christi area including some of those listed above. Nothing is known as to the fate of the holdings of this lab, which was closed down by SWRI around 1975. The data search by the present project was stonewalled by the SWRI San Antonio headquarters. Fortunately, a few individuals in the Corpus Christi area recognize the value of older data. One is Mr. John Buckner of the Coastal Bend Council of Governments, to whom the SWRI Ocean Science lab provided copies of their reports and some of their holdings. Mr. Buckner has diligently preserved this information for the past two decades and made his holdings available to the present project, from which SWRI routine data (Projects 5 and 13, Table 5-1) were keyboarded.

Another instance is the Corpus Christi field office of U.S. Geological Survey, which operated throughout the 1970's, collecting sediment chemistry data and hydrographic data (including circulation studies), but then was closed by USGS around 1980 and the personnel scattered. What holdings of this office survived are in the possession of Dr. Charles Holmes of the Denver USGS office, and appear to be only a fraction of the data actually collected by this office over the years.

4. Agency instability, i.e. dissolution, merging, reorganization, displacement & relocation.

Some data sets have survived by dint of being undisturbed. With an office move, as parcels, files and boxes are shifted about, the exposure to loss or discard is greatly increased. The disarray and haste usually typifying such moves contribute to a "clean-the-house" mentality, exacerbated by snap judgements on the part of personnel in no position to appraise the value of information. The decision is forced to consider data sets whose retention is already tenuous. Clearly, any sort of instability that leads to such shifting of materiel increases the probability of data loss.

5. Natural calamities (fires, floods, hurricanes) in poorly protected housing.

This problem speaks for itself. We have had a surprisingly large number of losses to such events on the Texas coast. Ironically, it is the large, centralized, difficult-to-duplicate sets that are most exposed. The usual problems of water leakage, faulty wiring, and deterioration operate everywhere, but the Texas coastal zone is exposed to extraordinary hazards. The human tendency is to disregard the risk of extreme hazard.

6. Changes in data management technology, without upgrading of historical files.

This is a surprising factor, at least to these authors. There are several forms of this technological hazard. The first is simple technological obsolescence. At the time of data entry, punched cards and 8-track formats seemed to be fixed technology. Now, they are virtually unreadable. There is a transition period, of course, when newer technologies replace the old, but the task of upgrading formats of large, rarely used data files is onerous and of low priority. Then, with the same apparent suddenness of the demise of the LP and the Magcard, the technological hardware support is no longer available. At this writing, many data sets which were "stored" on floppy disks are becoming as unreadable as 8-inch floppies.

A second variety of this hazard is software obsolescence, in which the encoding is no longer readable. This ranges from discontinuation of a proprietary software, to loss of the description of coding formats. The prominent example of the former is System-2000 data bases. There are several examples of the latter, in which there exist tapes containing numerical data which can be read but whose meaning is no longer documented. (We suspect one of these may be the reason that SWRI could not produce any data from its longtime study from the early 1970's.) In recent years, the proliferation of various "data-base managers" and "spreadsheet" softwares, combined with the increasing numbers of DOS and UNIX platforms, has greatly aggravated the problem, because most of these products are cross-platform and downward incompatible.

The third form of this hazard is due to the increasing information density of digital storage. As large data bases are compressed into smaller physical dimensions, the possibility of physical loss is increased: an errant electromagnetic field, small fire, or simple mislaying can wipe out the equivalent of reams of data. Probably the most prevalent form of this hazard is the acquisition of parity errors on an archival tape, and data garbling by stray magnetic fields. As new high-density media begin to appear, e.g. the Bernoulli box, the possibility of simple physical loss becomes greater. The key to minimizing this hazard is to create multiple copies. Technology is assisting in this, in that the cost for production of tapes and CD's is continuing to drop.

7. Proprietary attitude toward data by individual PI's.

This has been an endemic problem in academia, but it is also too frequently manifested in federal and state agencies. We will not propose to analyze the causes of this mentality, which may be rooted in the publish-or-perish environment, the paranoia of being "scooped" in some great insight gleaned from data analysis, the notion that "information is power," the view that one's data is valuable, and the view that one's data is worthless. We will observe that we did not encounter nearly as many cases of overt resistance in the Corpus Christi data search as in the Galveston Bay area. Most of the principal investigators and agency personnel were cooperative (and a lack of cooperation being due more to limits of time than intransigence). Perhaps this is because the problems in the Galveston Bay work have been widely publicized and corrective administrative actions have been taken. We are inclined to believe otherwise, that this is due instead to a basic cultural change presently underway, driven primarily by large-scale networking, *viz*. the Internet, which is fostering a more communal view of data resources.