

## 2. DATA SETS AND DATA PROCESSING

### 2.1 Parameters, their measurement and relations

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 parameters, some of which are indicator variables, such as coliforms in water, some of which are constituents which *per se* have major roles in biochemical processes, such as nutrients and pesticides, and some of which serve in both capacities, such as salinity. Some parameters are routinely measured in the field, such as temperature and salinity, and therefore the data base is most extensive for these variables. Most parameters are determined by laboratory analysis of a water, sediment or tissue sample. 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.

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 the study area may employ different measures, depending upon objective, convenience and tradition. The relations between parameters are important to the present analysis, for two reasons. First, from an analytical viewpoint, 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, and the creation of proxy data sets is treated here as an element of data processing.

#### 2.1.1 Temperature and salinity

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 created by the influx of freshwater.

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. Since there are large spatial gradients in salinity and it exhibits high temporal variability, a lower degree of precision in salinity determination can be accepted for work in estuaries than the case either in totally fresh or oceanic systems.

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). This gives salinity as a linear function of chlorinity and provides the means to convert from one to another (Defant, 1961, Wallace, 1974). One of the most common methods of salinity measurement is conductivity, particularly convenient for field determination since field conductivity meters are relatively inexpensive and widely available. One additional 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). Density (and specific gravity) is dominated by changes in salinity, and can serve as an alternative measure of salinity (though no such data were encountered in this compilation). The relationship of density to salinity and temperature, the equation of state for seawater, is presented in standard references (UNESCO, 1981).

Generally, in the field data from Corpus Christi Bay, one of the above measures is employed for determination of salinity, so the only decision available in analyzing the data is the proper conversion. On occasion, more than one method is used so there is a choice. This provides an opportunity to determine consistency and probable errors in measurement. Laboratory titrations and conductivity determinations, both field and laboratory, are occasionally presented for the same sample in the TNRCC SMN data base. To clarify the variability and relation among these different parameters, we analyzed those data records in which all three variables were measured. Widespread discrepancy was found, partly due to degraded accuracy in the laboratory determinations, and partly to the fact that a significant proportion of the reported laboratory values are not really measurements, but are "substitute data," apparently resulting from "rules-of-thumb" data entries instead of actual measurements. We refer to these as data Based On Graphical or Arithmetical Suppositions (BOGAS data). To summarize a detailed evaluation, the order of (decreasing) reliability was determined to be: field conductivity, laboratory chlorides and (lastly) laboratory conductivity. Thus, laboratory conductivity data were used only when other measures were unavailable, which resulted essentially in expunging the BOGAS data from the data base. Similar problems were encountered with the Texas Department of Health data base. Details are given in Ward and Armstrong (1997a, 1997b).

This experience with salinity data exemplifies two major points concerning data collection. 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. What then of the many programs in which only a single measure of salinity was made, and there is no means of cross-checking the data? 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. 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, especially the laboratory. 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 by analysis of the salinity data. What then can be expected of more complex and demanding analyses of trace metals or organics?

### 2.1.2 Dissolved oxygen

As noted above, 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 pitching 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 and 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, and low DO's are traditionally linked to the presence of oxygen-demanding pollutants.

One of the key controls on the concentration of DO is its solubility, which is a strong function of temperature and salinity and therefore varies substantially over the year. As temperatures range from perhaps 5 to 35 C and chlorinity from 0 to in excess of 45 ‰, 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, an associated parameter is defined, the oxygen *deficit*

$$D = C_s - C$$

where C is DO concentration and  $C_s$  is the solubility concentration, both in mg/L. The use of deficit 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 both parameters. Deficit, by itself, cannot be

interpreted biologically: a deficit of a given magnitude may be biologically limiting in summer and biologically unimportant in winter.

### *2.1.3 Suspended solids and turbidity*

Solid particles of greater density than water but which are small enough to be carried, even 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 over-emphasized, 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, erosion of the shoreline, and by transport by wind.

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 surrogate measure of suspended solids.

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. 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). 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. This numerical equivalence holds only for the calibration compound. For different types and distributions of suspended matter, NTU's and JTU's depart widely. 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, and their relationship is based upon a complex of theory and empiricism, see Preisendorfer (1986) and Effler (1988). Generally, there is reason to expect an inverse relation between turbidity and Secchi depth. From scattering theory and various empirical results, turbidity is found to be

roughly proportional to suspended solids (Jones and Willis, 1956, Di Toro, 1978). These relations were combined and calibrated with data from Galveston Bay (Ward and Armstrong, 1992) and Corpus Christi Bay using paired measurements of Secchi depth, turbidity and TSS from the TNRCC SMN data base. Each is a noisy measurement, and their interrelation is an additional source of uncertainty. Nonetheless, these relationships provide a vehicle for constructing a long-term proxy data base of suspended solids from a variety of turbidity and physical measurements.

#### *2.1.4 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 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. Fundamentally, the BOD is the amount of dissolved oxygen (DO) consumed in a sample of water during some period of time, for which 5 days is conventional.

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, including the types and initial quantities of bacteria (the “seed”) present in the water and their respective growth, the concentrations and chemical characteristics of the organic constituents within the water, the constituents which act as an inhibitor or a stimulant for the bacterial metabolism, environmental parameters (notably pH and temperature), and other aerobic organisms in the water, especially phytoplankton and zooplankton. It is apparent that there is a multiplicity of factors that can affect the BOD in the water parcel, which 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. The phenomenon may be due to nonlinearity of the rate constant as a function of the dilution factor, as implied by the Monod equation for bacterial growth (e.g. Monod, 1949), which of course contradicts the basic assumption underlying the dilution approach, that the BOD depletion is simply proportional to dilution.

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. There are some 435 paired measurements of BOD and TOC in the data base for the Corpus Christi Bay system. Analysis of these (Ward and Armstrong, 1997a) demonstrated that there is little correlation, the explained variance (of TOC on BOD) being a meager 15.5%. In addition to BOD, parameters such as volatile suspended solids (for water) and volatile solids (for sediment), and oil & grease (for both water and sediment) are often used as indicators of labile organics.

### *2.1.5 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 especially associated with waste discharges, notably nitrogen and phosphorus. These nutrients 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 in the water phase, and Kjeldahl nitrogen (the sum of ammonia and organic) in the sediment phase. As the relative proportions of nitrogen 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 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. With both nitrogen and phosphorus measurements, it is also important to differentiate between the “dissolved” and “total” concentrations in a water sample, i.e. whether the water sample is filtered for suspended matter before subjected to chemical analysis. Phosphorus, in particular, is sorptive and has an affinity for fine-grained suspended sediments.

Silicon in its various forms is an important nutrient in estuaries, especially in supporting the silaceous cells of diatoms, an extremely significant component of the phytoplankton. Silicon primarily originates 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 and Montague (1996), but in Corpus Christi Bay the data base is not extensive enough to allow this use of the parameter.

Total organic carbon in both water and sediment 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.

Chlorophyll-*a* has become an increasingly common water quality parameter, serving as a measure of phytoplankton biomass, though the actual content of chlorophyll-*a* in organic mass is strongly dependent upon the algal species involved. A related parameter is the metabolic product phaeophytin, sometimes loosely referred to as “dead” chlorophyll. In the Corpus Christi data, 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 chlorophyll-*a* and phaeophytin are virtually uncorrelated.

#### *2.1.6 Trace elements and organic pollutants*

Trace elements, primarily heavy metals, and hydrophobic organics (a.k.a. micropollutants, organo-xenobiotics, volatile organics), including pesticides, are more recent arrivals on the analytical scene. 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 known standards, in the process of which the reliability of the method as a function of standard concentration is quantified. These constituents occur in minute quantities, but many have disproportionate potential biological effects.

By far, the most common method for elemental analysis used for the measurements in the CCBNEP data base is spectrochemical, mainly atomic absorption (AA) and atomic emission (AE) spectrometry. 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 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 and a carrier gas through a chromatograph column, a glass tube coated with thin liquid film or packed with liquid-coated granular solids. 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 identification of the “chromatogram,” i.e. the pattern of signal (proportional to concentration) versus retention time, is accomplished by injecting standards of known composition. A more versatile operation is to couple the output from the chromatograph column into a mass spectrometer, referred to as GC-MS, to produce a so-called “mass spectrum.” Quantification again requires a calibration standard, and identifying mass spectra requires access to reference spectra, which can be facilitated by computer-automation of the GC-MS procedure. 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; (2) generally, the GC-MS output is searched for specified compounds, using 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.” It should be noted that the AAS, AES, and GC-MS techniques in fact produce a *number* for a specified 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 analytical procedures.

For the CCBNEP data base, 48 organic compounds and 31 metals for the water phase and 58 organic compounds and 17 metals for the sediment phase were identified for specific compilation and analysis. The organics may be broadly categorized as petroleum-derived hydrocarbons and pesticides. 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 the data compilation. Even at this, the data base for many of the compounds proved to be too sparse to support meaningful statistical analysis.

Especially for trace organics, protocols and procedures are still evolving, and this is reflected in a confusion of data acquisition and reporting. This is compounded by 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 (hexachloro-cyclohexane), as well as commercial names such as Lindafor, Lindagram and Lindamul (EPA, 1988). 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), whose use was banned in the U.S. in 1972, is certainly the most prominent of the organochlorines and the one for which the available data base in Corpus Christi Bay is greatest. 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 consistent with the current rule-of-thumb of a 3:1 ratio of p,p’-DDT to o,p’-DDT. Many agencies report simply total DDT. Usually this means the total for all isomers. The above proportion can be used as the relation between total DDT and p,p’-DDT, a workable proxy relation, though there are no paired measurements by which it can be tested.

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 compounds the analysis and identification problem. 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.

Polychlorinated biphenyls (PCB’s) is a class of biphenyl compounds containing up to 10 chlorine atoms (10 being the maximum number of sites available). 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). 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 ubiquitous pollutants. This is due to their widespread use,

but also to their mobility. Like chlorinated pesticides and many other organics, they are hydrophobic-i.e., have a propensity to sorb to 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.

Another important class of trace organics is polycyclic aromatic hydrocarbons (PAH's). Probably the most important physicochemical aspect of a PAH is its molecular weight. The light end of the range includes naphthalene ( $C_{10}H_8$ ), fluorene and anthracene, and the heaviest, benzo(a)pyrene, chrysene and coronene ( $C_{24}H_{12}$ ). 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 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. Similarly, with PAH's the "total PAH" is the arithmetic sum of the compounds analyzed. Some of the data sources for Corpus Christi Bay analyze only 4-6 PAH's, while others determine an extensive suite. (One data source reported two different "total PAH" values for the same water sample, analyzed by separate laboratories, which often differed substantially.) Parameters for total PAH and total PCB are included in the CCBNEP compilation. However, data sources for these were limited to those that reasonably approximated the same analytes.

### *2.1.7 Coliforms*

The reader has no doubt noticed that this section has addressed water and sediment quality 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. 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. There is, due to the regional shellfish industry, especially in the upper bays (Corpus Christi, Aransas, Copano) a considerable data set for the study area. Generally only one of total or fecal coliforms is measured.

To the extent that both are dominated by an origin in discharge of sewage, one might expect there to be a stable relation between the two. In fact, a rule-of-thumb is about, that fecal coliforms are approximately one-fifth of total coliforms (e.g., Kenner, 1978) but there seems to be little published support. Using the Galveston Bay data set, which included nearly 2000 paired measurements of total and fecal coliforms, Ward and Armstrong (1992a) found that the scatter in the data was too large to attach any significance to the mean of the fecal:total ratio. In fact, for practical purposes, the ratio approximated a uniform random distribution over the range 0 to 1. The conclusion that there is no useful ratio by which fecal and total coliform may be related was taken to apply to the Corpus Christi data set as well, and the two are treated as independent parameters.

## **2.2 Data sources for Corpus Christi Bay**

The data analyzed in this project were drawn from thirty past programs in Corpus Christi Bay, summarized in Table 2-1. Each of these comprises measurement of some of the water, sediment or tissue quality variables within a part of the Corpus Christi Bay study area for some definite sampling interval and period. Apart from this general feature, the programs differ in objectives and procedures.

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 presently 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 (TNRCC) 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, usually carried out by headquarters and/or the regional office of the Commission. Generally, field parameters are obtained *in situ*, by means of electrometric probes or portable analytical kits, and water/sediment

Table 2-1  
Current and historical sampling programs in Corpus Christi Bay study area  
providing data for CCBNEP Status and Trends analysis

<i>Code</i>	<i>Abbreviation</i>	<i>Agency or source</i>	<i>Project or Program</i>	<i>Format of source</i>	<i>Comments</i>
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. Reformatted 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 early-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

Table 2-1 (continued)

<i>Code</i>	<i>Abbreviation</i>	<i>Agency or source</i>	<i>Project or Program</i>	<i>Format of source</i>	<i>Comments</i>
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 (continued)	hard copy	Keyboarded by this project.

Table 2-1 (continued)

<i>Code</i>	<i>Abbreviation</i>	<i>Agency or source</i>	<i>Project or Program</i>	<i>Format of source</i>	<i>Comments</i>
16	MSI-LM	Marine Science Inst.ÄWhitledge	Laguna Madre nutrients data	digital EXCEL	Minor reformatting required.
17	MSI-NB	Marine Science Inst.ÄWhitledge	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./ TAMUÐCC	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 2-1 (continued)

<i>Code</i>	<i>Abbreviation</i>	<i>Agency or source</i>	<i>Project or Program</i>	<i>Format of source</i>	<i>Comments</i>
25	NCDH	Corpus Christi D 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.

samples are shipped to various laboratories for analysis. Parameters have been expanded from conventional variables in the early 1970's to trace constituents, pesticides and priority pollutants in recent years.

The Texas Parks & Wildlife Department and its predecessor agencies, the Texas Game and Fish Commission and the Texas Game, Fish and Oyster Commission, have monitored the fishery resources of the Corpus Christi Bay system for many years, and in association with this obtains 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 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 seafood in Corpus Christi Bay, the Seafood Safety Division (née 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 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 the Corpus Christi Bay system, as listed in Table 2-1, 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 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 2-1 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 included in this compilation, there are several which are noteworthy. These include the Texas A&M "Baseline" program of the 1950's sponsored by Reynolds Metals, the data collections of the Corpus Christi field office of the U.S. Geological Survey which operated throughout the 1970's, and the USF&WS-sponsored study of the system in the late 1970's performed by HDR and North Texas State University. Most important is probably that 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 were performed by this lab under sponsorship of regional agencies and industries. When SWRI closed the lab, all of its data holdings were removed and probably destroyed. Fortunately, John Buckner of the Coastal Bend Council of Governments made available to this project his considerable archive of 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 focused entirely upon sediment and is the only data set extant which sampled the *entirety* of the Corpus Christi Bay study area 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).

With respect to tissue data, two fundamental problems were encountered. First, for all of the agencies that routinely acquire tissue data, that data is managed differently from the water and sediment chemistry data. Therefore, these data sets did not simply entail an augmented acquisition procedure, but required special handling different from that of the water or sediment data. Second, for a specific chemical parameter, there is a greater range in what is measured and how it is reported in the tissue phase compared to water or sediment (described in more detail in Section 2.4 and in Chapter 3 below), and there is a corresponding lack of consistency among agencies, sometimes within the same agency. This aggravated the compilation problems, and led to lack of intercomparability from data source to data source, and therefore a reduction in statistical inference power. The net effect on tissue data analysis was that the information return on effort invested was disappointing.

The data programs of Table 2-1 formed the basis for the present analysis. Many of these programs, it will be noted, are small-scale research activities, though most of the data is dominated by the few large-scale programs summarized above. The approach of this project is 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, 1997b), along with any problems encountered in the data and how those problems were resolved (or reconciled). 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.

### **2.3 Segmentation of Corpus Christi Bay**

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. There are two broad

objectives for imposing a segmentation system on an estuary: administrative and analytical. The former refers to administration of laws and regulations. The latter refers to the aggregation and analysis of data of some sort from subregions of the bay, 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. 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. An example of a purely administrative segmentation is the state tract system, which is used by the General Land Office and several other state agencies, in which the constraints of operational surveying completely determine segment boundaries. It is a system devised for its administrative and operational benefits, rather than analysis of water quality.

One of the earliest, and therefore best-known, approaches to segmentation of an estuary for the purpose of water-quality analysis is that of Ketchum (1951a,b), who subdivided an estuary into segments of length equal to the tidal excursion. His segmentation is hydrographic in principle, 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 requirements on the segmentation. The most prominent example of segmentation for computational purposes is the gridding of a numerical model, such as the finite-element model of the Texas Water Development Board (Longley, 1994).

The most important administrative water-quality segmentation system in the study area is, of course, that of the Texas Natural Resource Conservation Commission. The Corpus Christi Bay system, including the tidal tributaries and the adjacent Gulf of Mexico, is presently subdivided into sixteen (16) segments, Table 2-2. T\_\_ TNRCC WQ 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 regulation, 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 TNRCC WQ segmentation is a combination of geography, tradition and politics. In this project, status-and-trends analyses were carried out for each of these TNRCC Water Quality Segments, and the study area was extended to include three classified freshwater segments, seven unclassified (both fresh and tidal), and one comprising a portion of the next bay system (San Antonio Bay). These results are presented in the appendices to the Final Report (Ward and Armstrong, 1997a).

Table 2-2  
TNRCC Water Quality Segments for CCBNEP Study Area

---

<i>Estuarine &amp; marine segments</i>	
2001	Mission River Tidal
2003	Aransas River Tidal
2101	Nueces River Tidal
2203*	Petronila Creek Tidal
2462*	Southwest 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
<i>Freshwater segments</i>	
2002*	Mission River above tidal
2004*	Aransas River above tidal
2204*	Petronila Creek above tidal
<i>Unclassified segments, estuarine &amp; freshwater</i>	
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

---

To secure the objectives of this project, it was necessary to perform analyses on a finer spatial scale than possible with the TNRCC segments. A system of “Hydrographic Segmentation” was devised for Corpus Christi Bay to form the basis for detailed analysis. The development of this segmentation begins with the observation that 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. Moreover, it is desirable to aggregate data from several sampling stations in order to create a sufficiently extended and dense set of data to allow statistical characterization of these specific water quality regions. Aggregation of data-i.e.,

segmentation-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 corresponds to the interior regions of segments and the latter to boundaries between segments. This should take into account transports, bathymetry, waste sources (where appropriate), inflows, and in general the distribution of physicochemical factors 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).

These notions were formalized and used as specific criteria of segmentation to guide the specification of analytical segments for the Corpus Christi Bay study area. 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, specification of water quality segments must include morphology and hydrography, hence our reference to this as "hydrographic segmentation." The segmentation system adopted for these analyses is depicted in Figs. 2-1 through 2-6. There is a total of 178 hydrographic segments, distributed as indicated in Table 2-3.

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, for example, it is appropriate to depict Oso Bay as two or three segments. Another researcher with the different purpose of studying the kinetics of a constituent within Oso Bay itself would find this scale of representation much too coarse, and would employ a much more refined spatial segmentation. Either level of

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

<i>system</i>	<i>segments</i>
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	

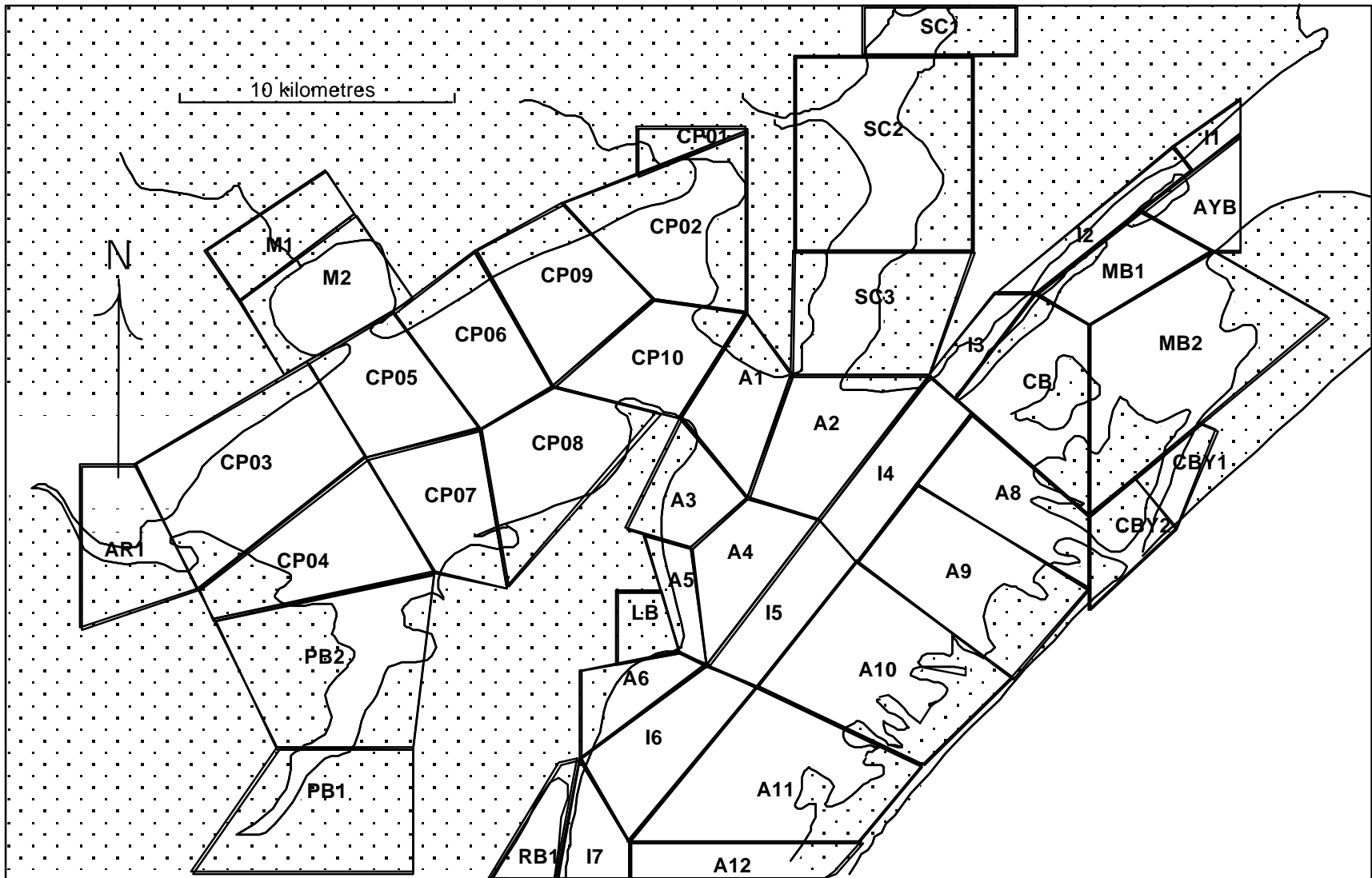


Figure 2-1. Hydrographic areas for Aransas-Copano system

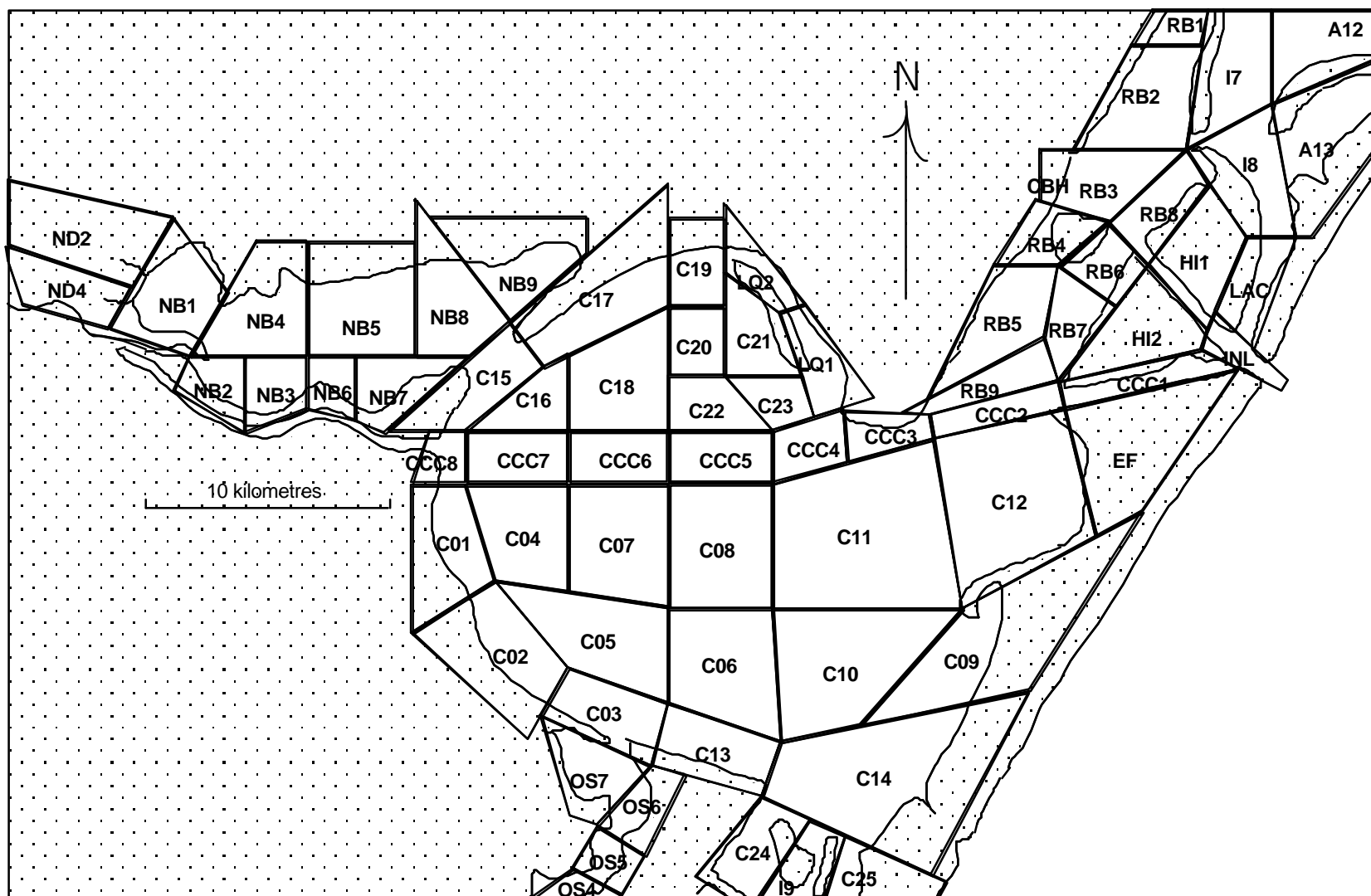


Figure 2-2. Hydrographic areas for Corpus Christi Bay

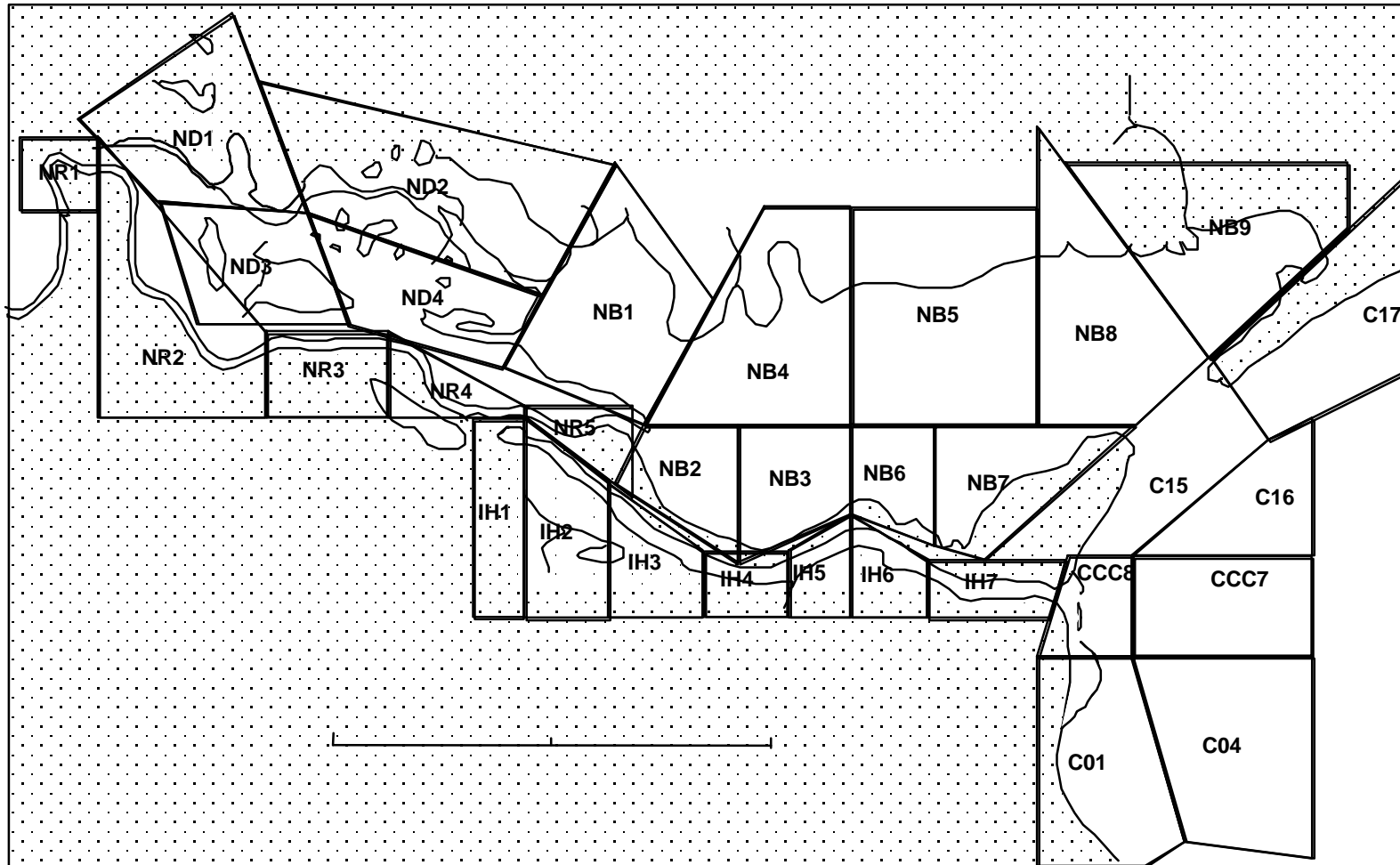


Figure 2-3. Hydrographic areas for Nueces Bay region

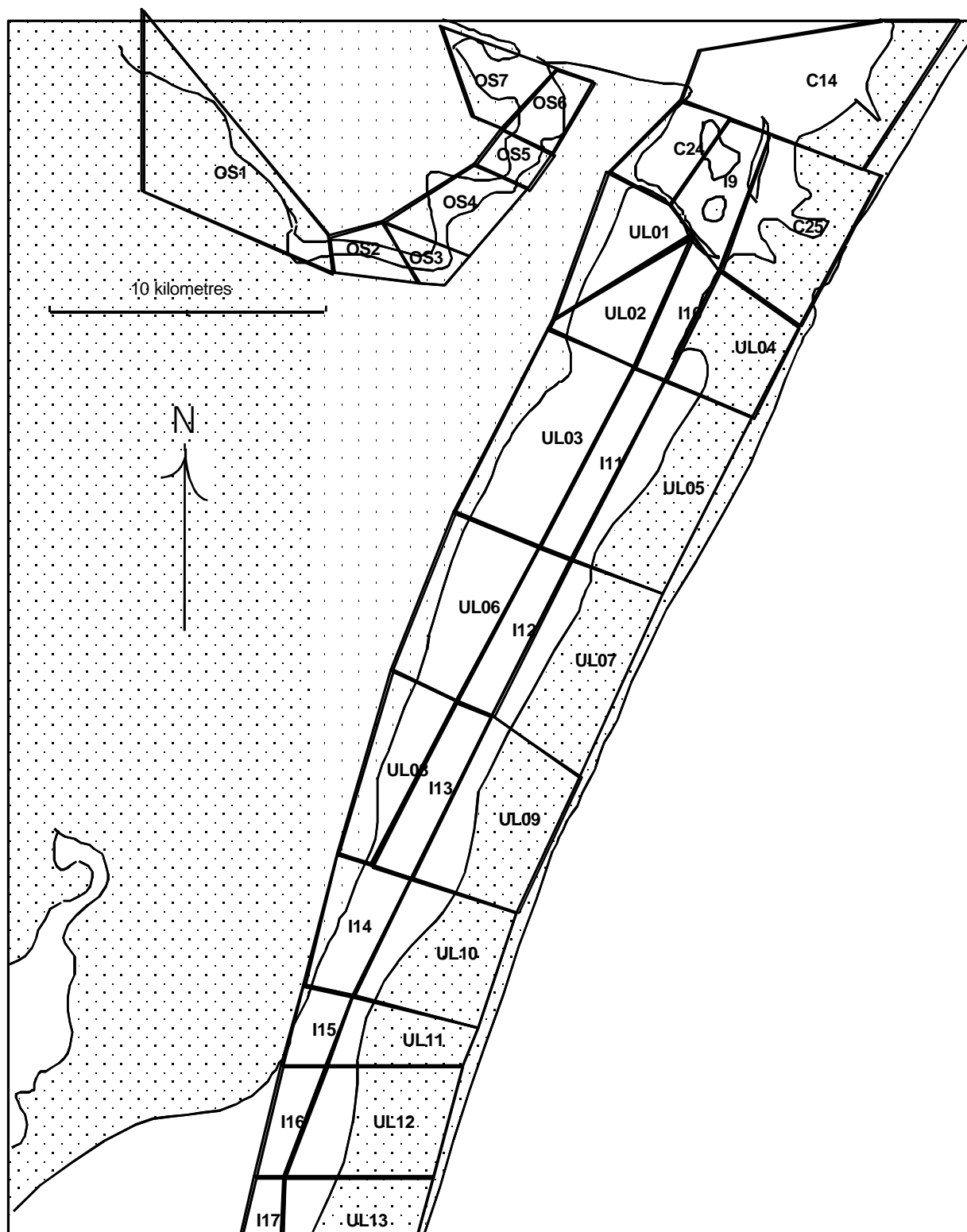


Figure 2-4. Hydrographic areas for Upper Laguna Madre and Oso Bay



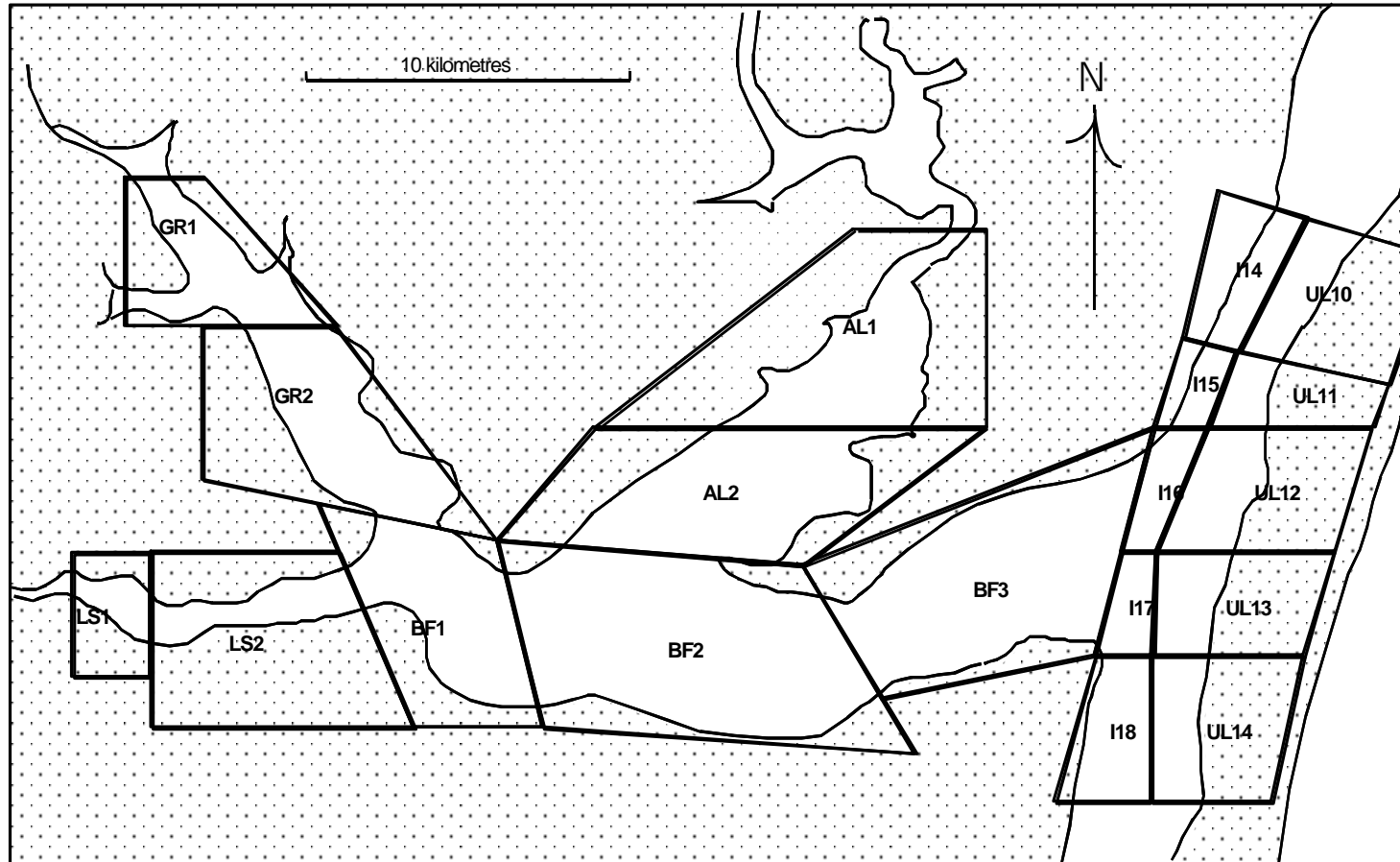


Figure 2-5. Hydrographic areas for Baffin Bay and adjacent Laguna Madre

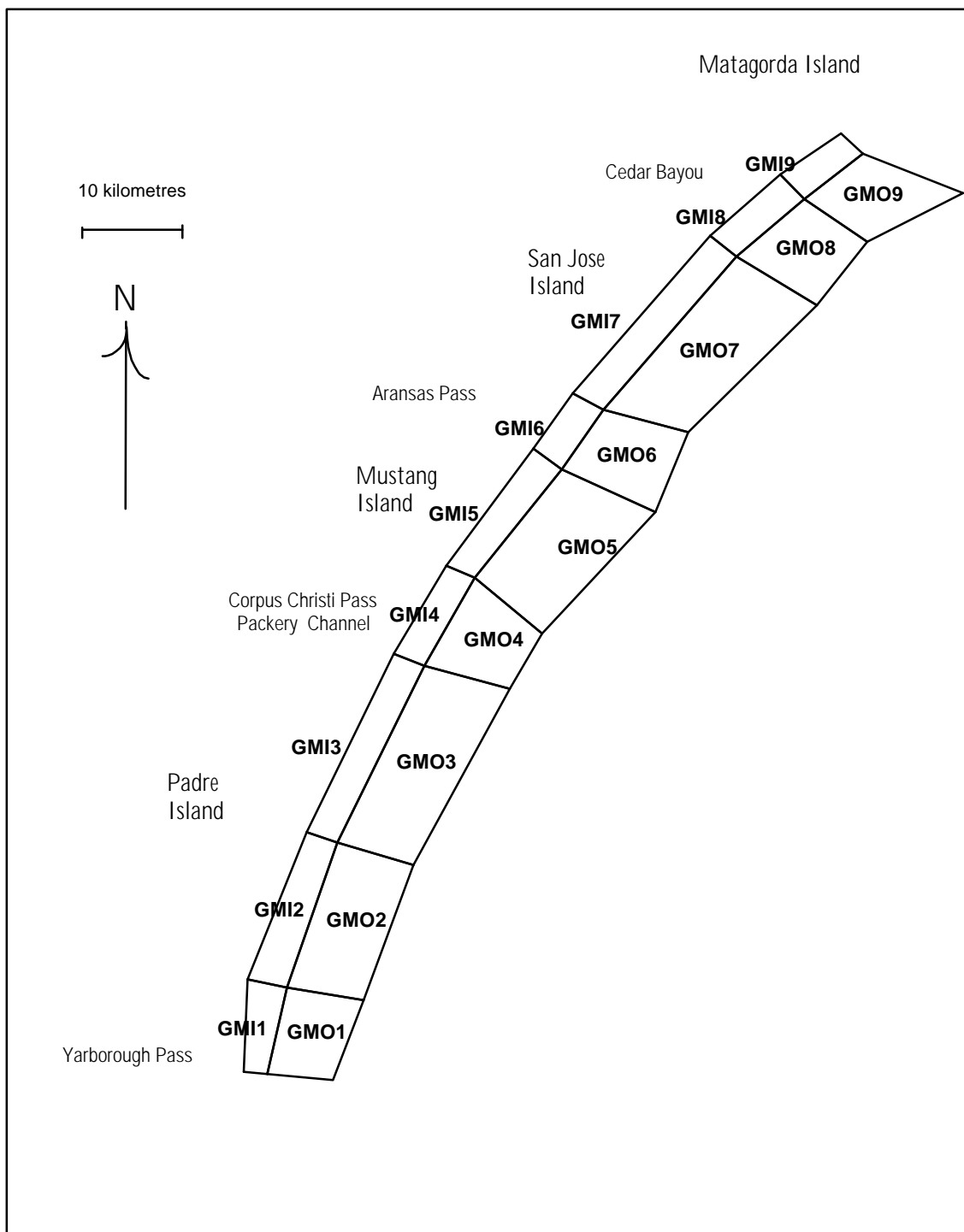


Figure 2-6. Hydrographic areas for Gulf of Mexico nearshore

segmentation would be inappropriate and unworkable for the other's purpose. (Note that the specification of a network of sampling stations implicitly assumes a spatial scale, in that each sampling station is presumed to represent water quality over some extended area in which the station is located.)

For data processing purposes, each segment was defined as the union of quadrilaterals encompassing the watercourse lying within that segment. The corners of each quadrilateral are given by latitude/longitude pairs. 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 quadrilaterals for both the Texas Natural Resource Conservation Commission Water Quality Segments and the hydrographic-area segments of this project are given in the appendix to the final report (Ward and Armstrong, 1997a).

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. There are two prominent breaks in the slope of the seabed, one at approximately 5 fathoms depth (10 m) and one at approximately 10 fathoms (20 m). Hydrographic segments were defined to extend between these two breakpoints. 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 refraction zone of the longer-length waves. 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.

The differences between the TNRCC Water Quality and CCBNEP hydrographic-area segmentations are:

- (1) The TNRCC segments tend to be larger in space, especially within the open bays, and may have arbitrary or political boundaries;
- (2) The CCBNEP hydrographic segments are smaller in spatial extent and are defined by principal geomorphic controls on flow and/or known predominant flow patterns;
- (3) The TNRCC segments include tributaries (tidal and fresh) of the principal inflows;
- (4) The hydrographic segments focus upon the bay system *per se*, its immediate periphery, and the nearshore Gulf of Mexico, but do not consider the upper reaches of the tributaries;

Analysis by TNRCC segments has the advantages of: (a) treating a smaller number of segments, (b) corresponding to the administrative framework for the Corpus Christi Bay study area, and therefore allowing direct comparison with standards and past surveys. On the other hand, analysis by CCBNEP hydrographic-area segments allows: (a) better spatial definition of variability, especially in the open bays, and (b) more realistic definition of areas that are nearly homogeneous, i.e., based upon hydrographic controls rather than political boundaries or convenience of access, therefore greater precision in the analyzed data.

## **2.4 Data base compilation**

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. All of the hard-copy data were keyboarded as a part of this project effort. 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.

### **2.4.1 Data Set Construction**

Data-base formats were devised specific to this project. To facilitate transfer and use of the data by other workers, emphasis was placed on data structure that is manipulable via

microcomputers (especially PC's), i.e. compact ASCII files. 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 (1997b), which is intended to serve also as a User's Guide to the data.

One of the principles observed in the construction of the Corpus Christi Bay data base was the maintenance of integrity of the original data. The *source* data files are distinguished from *derivative* data bases. The source data file codifies the original measurements as reported by the originating agency. It contains exactly the information in the original: nothing is lost or added. Even the original units of measurement are retained, since an apparently innocuous conversion of units can introduce a distortion. 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. For various analytical purposes, the 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 and to subject these to specific processing. There is a derivative file for each water/sediment-quality parameter of concern, comprised of records of measurements at points in time and space. Each data record in the derivative files also includes coded information identifying the data source, say TNRCC Statewide Monitoring Program, versus Corps of Engineers, versus TWDB Bays and Estuaries Program.

The complete list of water and sediment parameters for which data bases were compiled is given in Table 2-4. To facilitate reference to these parameters and to coordinate with file names and graphics, a unique abbreviation is employed for each parameter. 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 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

Table 2-4  
Abbreviations and units for CCBNEP water and sediment parameters

<i>abbreviation</i>	<i>definition</i>	<i>units</i>
<b>- water analytes -</b>		
WQALK	total alkalinity (as CaCO <sub>3</sub> )	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/100ml
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 <sub>4</sub> )	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 <sub>2</sub> )	mg/L
WQSO4	total sulfate (as SO <sub>4</sub> )	mg/L
WQTCOLI	total coliforms	MPN or colonies/100ml
WQTEMP	temperature	degrees C
WQTOC	total organic carbon	mg/L
WQTOTP	total phosphorus (as P)	mg/L
WQTPO4	total phosphate (as PO <sub>4</sub> )	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 2-4  
(continued)

<i>abbreviation</i>	<i>definition</i>	<i>units</i>
<b>- water analytes continued-</b>		
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	acenaphthene	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 2-4  
(continued)

<i>abbreviation</i>	<i>definition</i>	<i>units</i>
<b>- water analytes continued-</b>		
WQ-CHLR	total Chlordane	µg/L
WQ-CHLRC	Chlordane cis isomer	µg/L
WQ-CHRY	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	hexachlorobenzene	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	naphthalene	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-PHNA	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 2-4  
(continued)

<i>abbreviation</i>	<i>definition</i>	<i>units</i>
<b>- sediment analytes (dry weight)-</b>		
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 ignition)	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	acenaphthene	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 2-4  
(continued)

<i>abbreviation</i>	<i>definition</i>	<i>units</i>
<b>- sediment analytes continued-</b>		
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

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 acenaphthene in a water sample, and SED-LIND to the lindane in a sediment sample.

The situation is more complicated for tissue analytes. The tissue burden of a specific compound is determined by acquiring an organism from the estuary, excising a portion of that organism, mechanically homogenizing the excised portion, and performing a chemical analysis using generally the same methodologies as for a solids sample. Options are:

- |   |   |
|---|---|
| ● element or chemical compound<br>for analysis                                    | ● selection of organism   |
| ● organism portion to be analyzed:<br>whole organism<br>edible portions (“filet”) | ● reporting convention:<br>wet-weight concentration<br>dry-weight concentration |

Choice depends upon the ultimate purpose, which may be either (1) to determine flux of specific compounds or elements through the food chain; (2) to establish whether there is a public health risk entailed by consumption of that organism. All of this entails a great range of variation in the nature of the data. There is no *a priori* correspondence between whole-organism and filet concentrations (with the lone exception of the oyster, for which the two are equivalent). Without a specific determination of tissue moisture content, the dry-weight and wet-weight values cannot be interconverted. In the CCBNEP data set, tissue data was compiled for 100 analytes, namely metals and various trace organics (some of which were not analyzed in water or sediment samples), but these had to be further differentiated according to whether the concentrations were on a wet-weight or dry-weight basis, and whether the analyses were whole-organism or edible portions. This expanded the effective number of analyte/reporting combinations to 172.

The most fundamental parameter used to differentiate and classify tissue data is the species of the organism. The body burden of an organism is dependent upon the mechanics of ingestion (or absorption), details of metabolism, and the exposure time-line. These, in turn, depend upon the life history of the organism, its physiology, its regional habitat, and its position in the food web. Some, like many finfish and shrimp, move large distances through the system, as well as through the nearshore Gulf, while others are territorial, like the black drum, and still others are sessile, like oysters. The exposure to sources of contamination vary accordingly. Some, like filter feeders and some top carnivores, will exhibit a greater degree of concentration than others. Therefore, comparison of tissue data between species is almost never valid. In the CCBNEP data set, the sampled organisms were assigned a two-digit code, and this information was included in the data record (replacing depth in the record format structure). Table 2-5 lists the organisms encountered in the data compilation. Most of these were obtained opportunistically, and many occur only once in the data set, e.g. the gar and the

Table 2-5  
CODES FOR TISSUE ORGANISMS

Code	Common name	Specific name
00	unknown	no information provided
01	southern flounder	<i>Paralichthys lethostigma</i>
02	fin perch	unknown
03	speckled trout	
04	American oyster	<i>Crassostrea virginica</i>
05	hardhead catfish	<i>Arius felis</i>
06	gafftopsail catfish	<i>Bagre marinus</i>
07	Atlantic croaker	<i>Micropogonias undulatus</i>
08	brown shrimp	<i>Penaeus aztecus</i>
09	penaeid shrimp (undiff.)	<i>Penaeus spp.</i>
10	blue crab	<i>Callinectes sapidus</i>
11	toadfish	<i>Opsanus beta</i>
12	calico crab	<i>Eriphia gonagra</i>
13	shoalgrass	<i>Halodule wrightii</i>
14	sheepshead	<i>Archosargus probatocephalus</i>
15	black drum	<i>Pogonias cromis</i>
16	red drum (redfish)	<i>Sciaenops ocellatus</i>
17	clam	<i>Mercenaria</i>
18	menhaden	<i>Brevoortia patronus</i>
19	whiting	
20	white shrimp	<i>Penaeus setiferus</i>
21	sea catfish	
22	ladyfish	
23	alligator gar	
24	carp	<i>Cyprinus carpio</i>
25	pinfish	<i>Lagodon rhomboides</i>
26	tarpon	<i>Megalops atlantica</i>
27	spot croaker (spot)	<i>Leiostomus xanthurus</i>
28	mullet	<i>Mugil spp.</i>
29	stone crab	
30	Spanish mackerel	
31	pigfish	
32	longnose killifish	<i>Fundulus similis</i>
33	perch	unknown
34	spotted seatrout	<i>Cynoscion nebulosus</i>

toadfish. The reader may find it hard to believe that there might be a need for category 00. In fact some of the tissue data from Corpus Christi Bay provided date and location of catch of the organism, and the concentrations for the suite of analytes, but failed to include the identity of the organism.

The final piece of information in the data record is the position of the sample in space. For all samples, this is the horizontal position. For water samples, the vertical position is also needed. Establishing the horizontal position coordinates can be problematic. Most of the 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 the digital data base.

#### *2.4.2 Quality assurance and reliability*

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 need for determining the reliability of historical data and discounting measurements that are judged to be “unreliable” is clearly important. It is our 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. The former is addressed in this section, and the latter in Section 2.4.3, following.

The CCBNEP 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 employed to minimize the occurrence of error, and maximize its detection. All data available in machine-readable form from an originating agency were obtained, manipulated and entered in that form. Further, intermedia transfers were minimized, e.g., copies were sought as ASCII or WK1 files that could be downloaded by ftp via the Internet or transferred on diskettes. When keyboarding was necessary, data-entry procedures utilized simple software with formats that mimicked the source hard copy, and were implemented to minimize factors such as fatigue.

We note that several of these procedures are significant departures from those recommended by Tetra Tech (1987) for National Estuary Programs. For example, Tetra Tech (1987) recommends 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, as it reduces efficiency and magnifies the chances of error. 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 (1987) recommends.

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, spatial continuity and temporal continuity, 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 a typographical error was “patently obvious.”

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 plotted. Generally, finer corrections were reserved for the derivative data-base screening unless some independent information was available.

A new source of station location error encountered in this data compilation is due to increasing use of high-technology positioning systems. These systems are radio-navigation systems, for example LORAN (née Loran-C), which has been in use in marine operations for years, and Geographical (a.k.a. Geodetic, a.k.a. Global) Positioning Systems (a.k.a. Satellites), GPS, a relative newcomer. These are now being embraced by field sampling operations because of the recent availability of economically priced receivers and on-board processors. In this data compilation, we found their accuracy to be much poorer than one might expect, probably due to environmental factors that degraded the performance of the equipment, to failure of the crew using this equipment to appreciate the possibility of error, and possibly to operational errors on the part of the field crew. These positioning systems as a source of error are addressed in the technical report and in more detail in the Data Base Report (Ward and Armstrong, 1997a, 1997b).

#### *2.4.3 Uncertainty measures and data quality*

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. 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. Each entry also includes a measure of the degree of uncertainty. 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 to which the data may be subject.

This uncertainty measure is the magnitude of the population standard deviation about a fixed value of the variate. It is estimated by the standard deviation about the mean of a series of measurements performed under controlled laboratory conditions, as reported in the literature on methodology for that parameter. The uncertainty may vary with the magnitude of the measurement, and is often generalized as a linear function. This is the format used in the most recent USGS manual (Fishman and Friedman, 1989) for dissolved analytes (see also Friedman and Erdmann, 1982). It is also consistent with the October 1992 policy of the National Institute of Standards and Technology (NIST), on expressing measurement uncertainty, based upon recommendations of the International Committee for Weights and Measures (1992). 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 utilized these uncertainties, 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) QA/QC procedures documented by the collecting agency, (b) literature estimates for the specific methodologies used, (c) statistical variation of the measurements themselves, relative to some external standard, (d) judgement of the principal investigators, based upon experience with the method or equipment.

One element in the formulation of uncertainty that has special significance is the threshold value of concentration, which must be exceeded for the analysis to be meaningful. 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, the detection limit is approximately three times the standard deviation for blanks.

For trace concentration determinations based upon instrumental methods however, especially those of elemental or trace organic 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 three standard deviations has a 1% chance of being in fact blank (zero). For measured concentrations less than this level, the probability is greater than 1% 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, and their use has been widely 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. 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).

Generally, there is more information-and more quantitative scope-on precision in the later literature than the earlier, which raises a dilemma: when precision information changes, should we utilize the data contemporaneous with the measurements, i.e. assumed to be reflective of the technology and procedures of the time, or should we presume that the more recent data derives from a larger base of measurements, and represents 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 we have elected the former. This is not an irreversible decision, as any later user of the data base has the option of employing a different measure of precision, and consequently a different data rejection procedure. 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].”

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 our 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 was applied *at the earliest* to the compilation of the derivative data files, not to the source data.

Data rejection can be performed based upon either the level of uncertainty of the measurement or its magnitude relative to a rejection trigger. 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.”

#### 2.4.4 Data processing

The principal steps in data processing in this study for all phases, i.e. water, sediment or tissue, were:

- (1) For each parameter (and organism) of concern, sift through the Source Data Files, applying whatever screening, proxy relationships, and units conversions are necessary, and order chronologically to create a Master Derivative File for that parameter;
- (2) Sort the Master Derivative Files into TNRCC Water Quality or CCBNEP hydrographic-area segments;
- (3) For each segment in either segmentation system, carry out statistical and/or graphical analyses of the data.

We regard the Master Derivative data files to be a chief product of this project, capable of much more study and analysis than are given here.

Generally, as a matter of personal philosophy, we rejected very little data in the formulation of the Derivative Files, reserving further data screening for the specific analyses to which the Derivative Data Bases are subject. Data were rejected from the Derivative Files if the date, position, or depth were obviously impossible and there were no satisfactory means of judging

the correct value. A minimum of screening was done based upon parameter value at this stage of the processing, mainly screening out clearly ridiculous values.

For many of the parameters, such as metals and pesticides, the lower range of measurement is delimited by the detection limits of the procedure, and detection limits are generally reported as part of the data set. Despite this, several of the data sets, especially those from the 1970's, include zero values. One source for these zeroes, we believe, is replacement of "censored" analytical reports, i.e., 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. Another possible source of the zeroes, especially in the older data from the era of punched cards, is that they may have originated from blank entries (i.e., no measurement) that in the process of agency transcription and digitization were replaced with a zero. Unfortunately, there is no foolproof means of distinguishing these. In the formulation of the derivative data bases, zeroes are allowed to stand. Later in the analysis phase, these are retained, deleted, or replaced by applicable detection limits, depending upon the specific parameter. For the analyses of trace organics and metals, 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, or reported by contemporaneous labs in other data sets. (The parameters affected and these estimated detection limits are tabulated in Table 4-2 of Ward and Armstrong, 1997a.)

The treatment of detection limits in analysis of water quality is particularly vexing. 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 above 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 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 to best suit the purpose at hand.

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 into multi-sample averages. Would one accept the existence of sea monsters as an extrapolation from large 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.

## **2.5 Summary of the data base**

The combined data base of all raw measurements, including proxy variables but excluding derived variables such as DO deficit and stratification, comprises approximately 480,000 individual point measurements, requiring about 24 megabytes of storage in a compact flat-ASCII format. About 43% of this data is made up of the hydrographic variables salinity, temperature, DO and pH. About 35% are conventional water-quality analytes, 14% are water-phase metals and 1% are water-phase organics. There is roughly an order of magnitude less sediment data from the Corpus Christi Bay system than water quality data, about 34,000 measurements. However, sediment transport processes and kinetics vary on time scales longer than that of the overlying water, so relative to the time and space scales of natural variability less data would be needed to characterize sediment chemistry than water chemistry. Of this data base, tissue data represents 8,200 separate measurements, primarily metals.

The data bases for water quality and for sediment quality are summarized on a systemwide basis in Tables 2-6 and 2-7. These tables 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

Table 2-6  
Summary statistics for water analytes for entire CCBNEP study area

<i>Parameter*</i>	<i>No.of obs</i>	<i>Avg &gt;DL</i>	<i>St dev &gt;DL</i>	<i>No. &gt; DLs</i>	<i>% &gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0      =DL</i>	
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	750827	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 2-6  
(continued)

Parameter	No.of obs	Avg >DL	St dev >DL	No. > DLs	% > DLs	Min	date	Min >0	date	Max	date	Average with BDL=0 =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 2-6  
(continued)

<i>Parameter</i>	<i>No.of obs</i>	<i>Avg &gt;DL</i>	<i>St dev &gt;DL</i>	<i>No. &gt; DLs</i>	<i>% &gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0 =DL</i>	
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
WQMETPBTD	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)													

(continued)

<i>Parameter</i>	<i>No.of obs</i>	<i>Avg &gt;DL</i>	<i>St dev &gt;DL</i>	<i>No. &gt; DLs</i>	<i>% &gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0 =DL</i>	
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-CHRY5	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 2-6  
(continued)

<i>Parameter</i>	<i>No.of obs</i>	<i>Avg &gt;DL</i>	<i>St dev &gt;DL</i>	<i>No. &gt; DLs</i>	<i>% &gt; DLs</i>	<i>Min &gt;0</i>	<i>date BDL=0</i>	<i>Min =DL</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with</i>	
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 2-4



Table 2-7  
Summary statistics for sediment analytes for entire CCBNEP study area

<i>Parameter*</i>	<i>No. of obs</i>	<i>Avg &gt;DL</i>	<i>Stdev &gt;DL</i>	<i>No.&gt; DLs</i>	<i>%&gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0 =DL</i>	
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	840823	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 2-7  
(continued)

<i>Parameter*</i>	<i>No. of obs</i>	<i>Avg &gt;DL</i>	<i>Stdev &gt;DL</i>	<i>No.&gt; DLs</i>	<i>%&gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0 =DL</i>	
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 2-7  
(continued)

<i>Parameter*</i>	<i>No. of obs</i>	<i>Avg &gt;DL</i>	<i>Stdev &gt;DL</i>	<i>No.&gt; DLs</i>	<i>%&gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0 =DL</i>	
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 2-7  
(continued)

<i>Parameter*</i>	<i>No. of obs</i>	<i>Avg &gt;DL</i>	<i>Stdev &gt;DL</i>	<i>No.&gt; DLs</i>	<i>%&gt; DLs</i>	<i>Min</i>	<i>date</i>	<i>Min &gt;0</i>	<i>date</i>	<i>Max</i>	<i>date</i>	<i>Average with BDL=0      =DL</i>	
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	910805	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	7.48
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	3.48	12	46	9	0	751103	10	830629	50	760831	0.322	41.1

\* Abbreviations & units defined in Table 2-4

these parameters, the baywide means have little importance; however they are useful in typifying the magnitudes of the different variables. Averages and standard deviations computed by the strategy of simply neglecting BDL measurements are given in the third and fourth columns of Tables 2-6 and 2-7. In the last two columns of these tables are given the averages computed by the strategies of assigning zero and DL values to the BDL's, respectively. Because of the frequent occurrence of zero values (rather than censored values), especially in the data from the 1970's, the minimum value exceeding zero is also tabulated in this summary. For some parameters, this may provide a better index to the range of the parameter.

It is important to note that these tables characterize the data *sets* rather than the *parameters*, because these data have not been screened for bad data other than the very liberal bounds on what would be considered "ridiculous." Indeed, the ranges of some of the variables disclose the presence of possibly spurious entries in the data set. These tables also illustrate the dilemma of applying rejection triggers. For example, 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. A few sediment volatile solids values exceeded 20%, and were rejected; though lying within the range of *possibility*, these seem improbable. (Even the value of 19% that remains in Table 2-7 is suspicious. Surely there would have been reports of lab technicians lacking eyebrows.) Below this, though, the demarcation between the reasonable and the unreasonable 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 a strictly *a priori* basis.

Many of the EPA priority pollutants do not appear in Table 2-4 or in Tables 2-6 or 2-7. Very few measurements of these have been made in the Corpus Christi Bay system. In some instances, there may be a scattering of measurements, but not enough to use in any meaningful way in a status-and-trends analysis. For example, there were only two measurements of water-phase Endrin aldehyde from the entire Corpus Christi Bay system. Most of the individual PAH's were represented only by a handful of data. Those variables for which the sample base is totally lacking or inadequate are excluded from these tables. Even 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.

Figures 2-7 through 2-28 display graphically the sampling intensity in regions of the study area for a few of the more important water and sediment parameters. Sampling intensity is measured by the number of observations within each hydrographic-area segment for the period of record. The amount of data available is strongly dependent upon the parameter. Further, what might appear to be a large number of historical samples for a given parameter on a baywide basis, from Tables 2-6 and 2-7, is shown to be quite modest-even inadequate-when related to specific areas of the bay. Sampling intensity is highly heterogeneous in

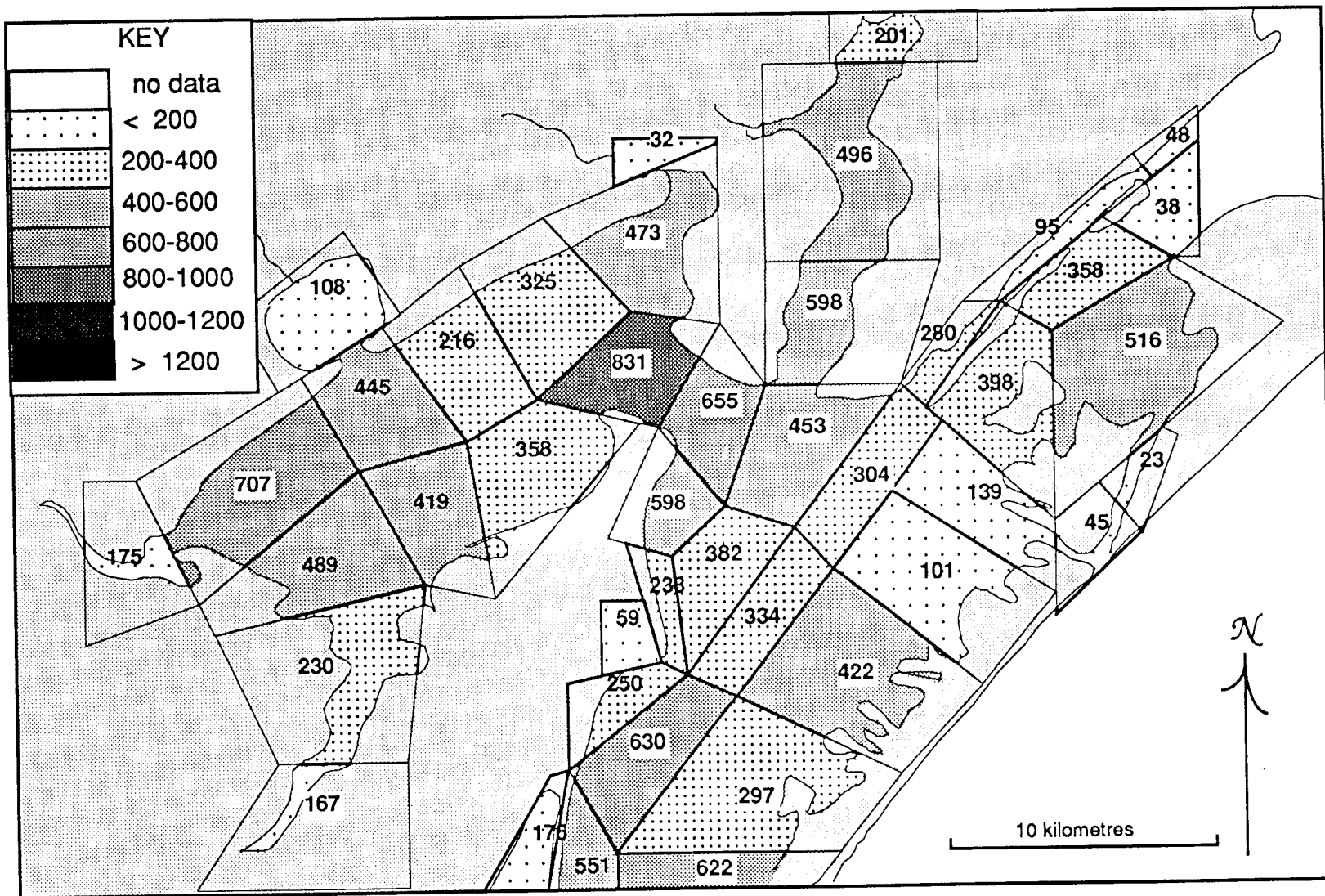


Figure 2-7. Sampling density (total number of measurements) of WQSAL for Aransas-Copano system

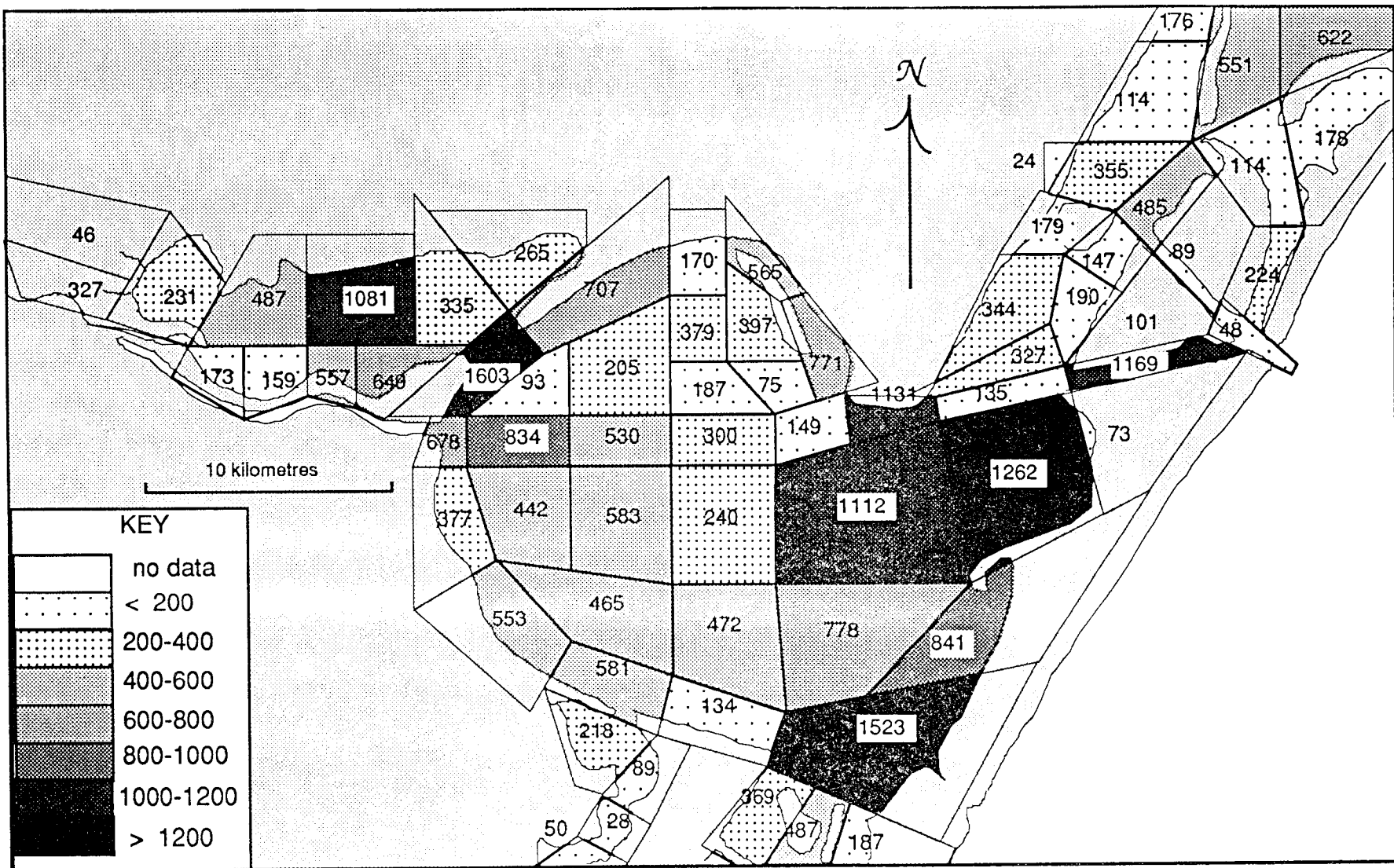


Figure 2-8. Sampling density (total number of measurements) of WQSAL for Corpus Christi system

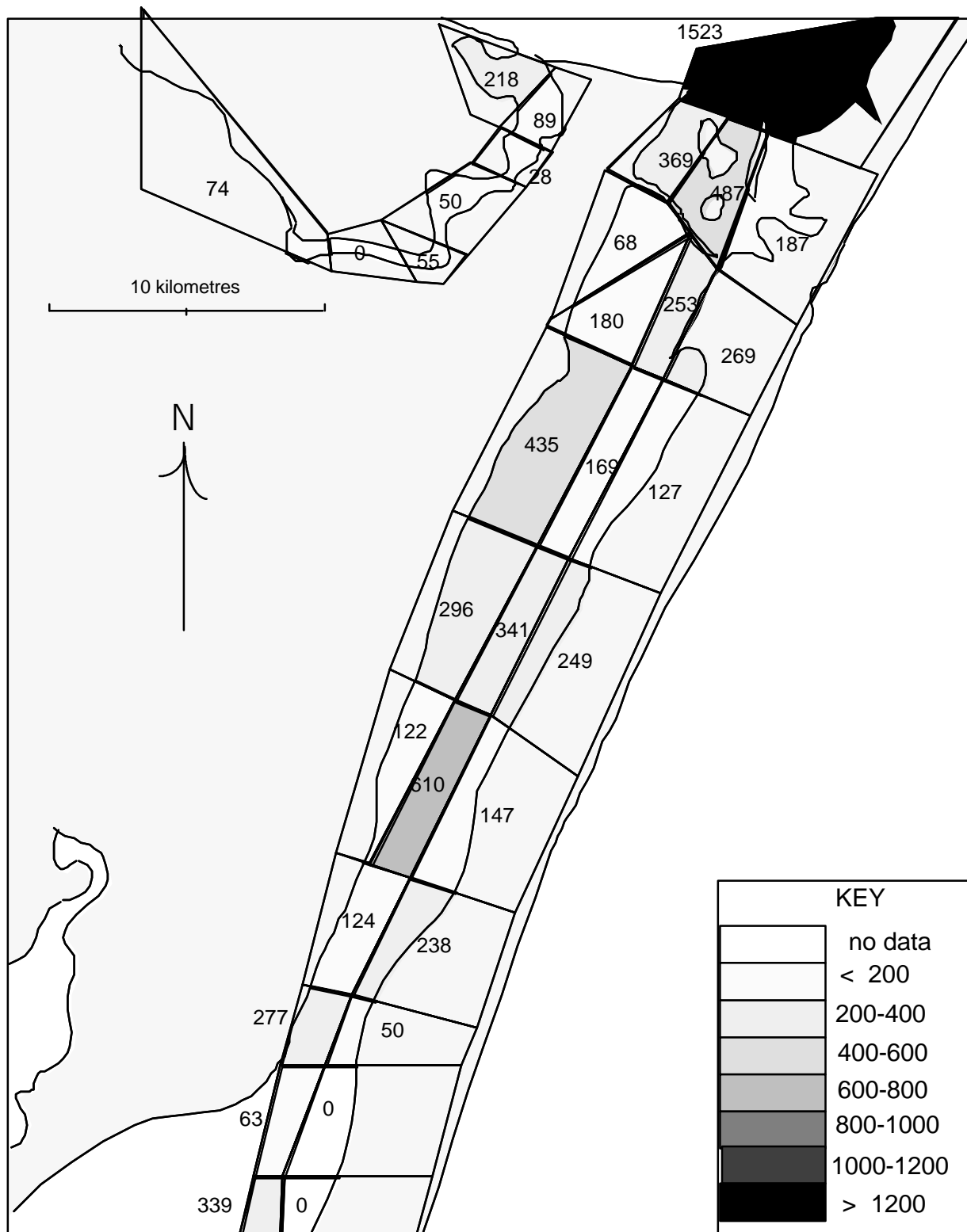


Figure 2-9. Sampling density of WQSAL for Upper Laguna Madre



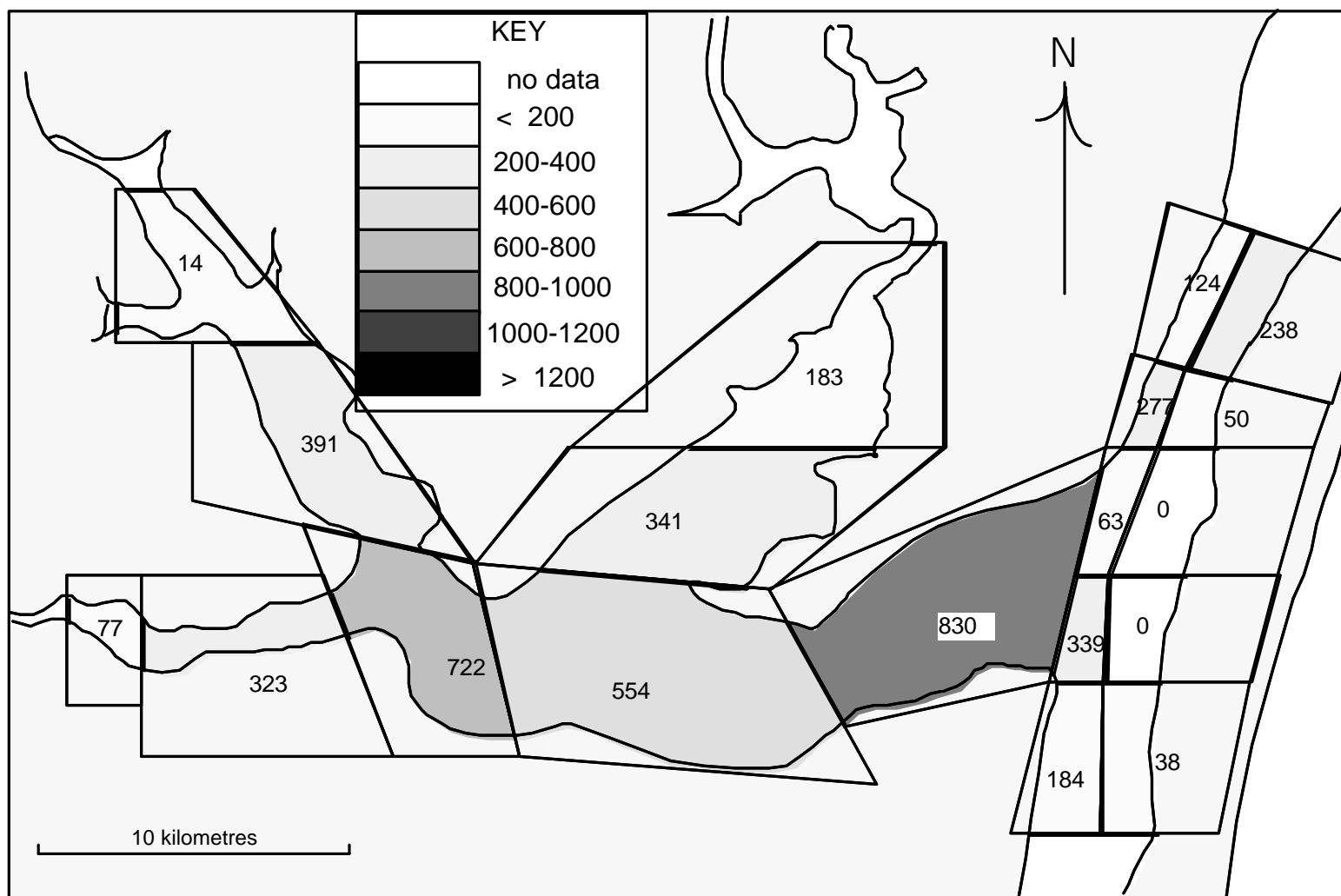


Figure 2-10. Sampling density (total number of measurements) of WQSAL for Baffin Bay system

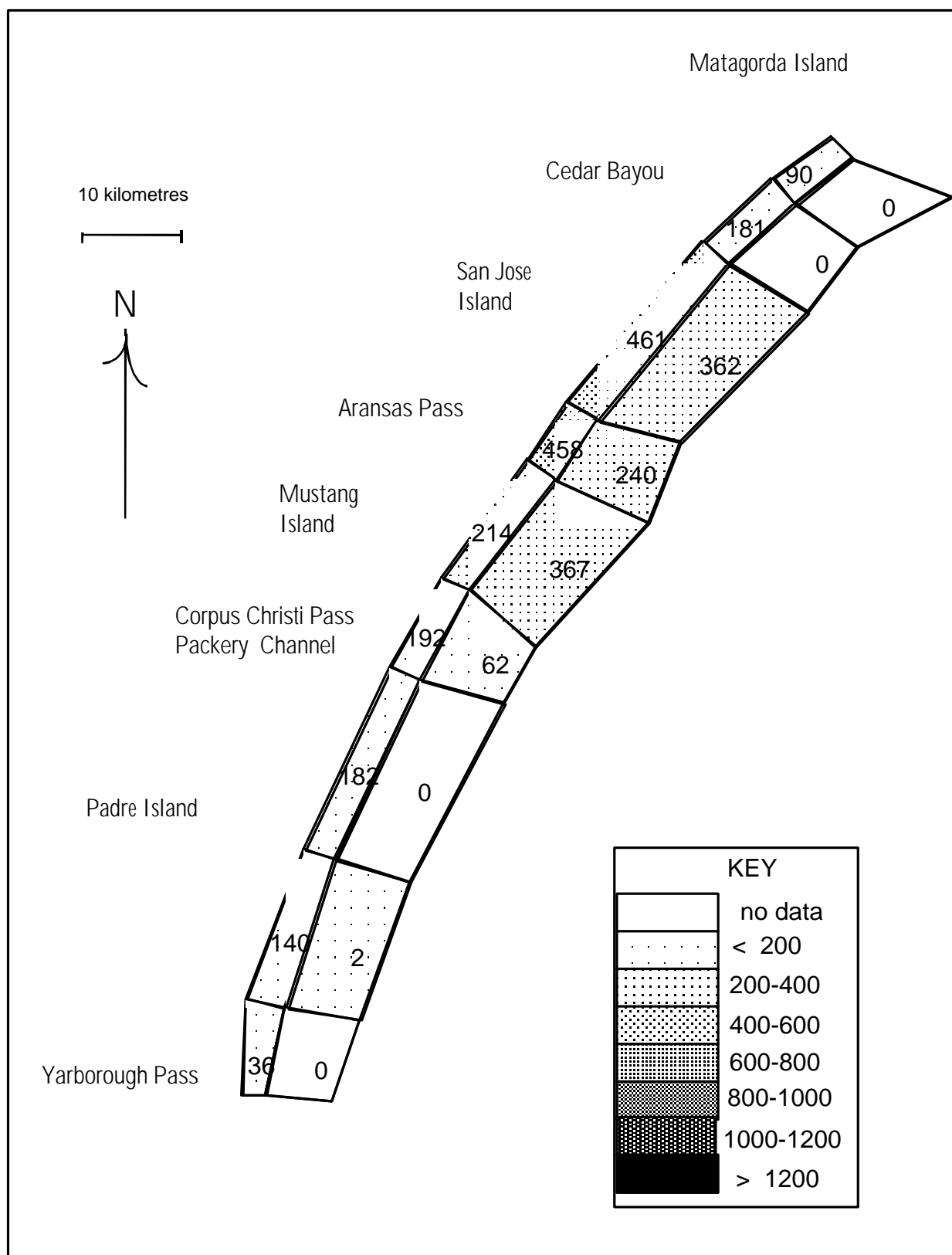


Figure 2-11. Sampling density of WQSAL for Gulf of Mexico



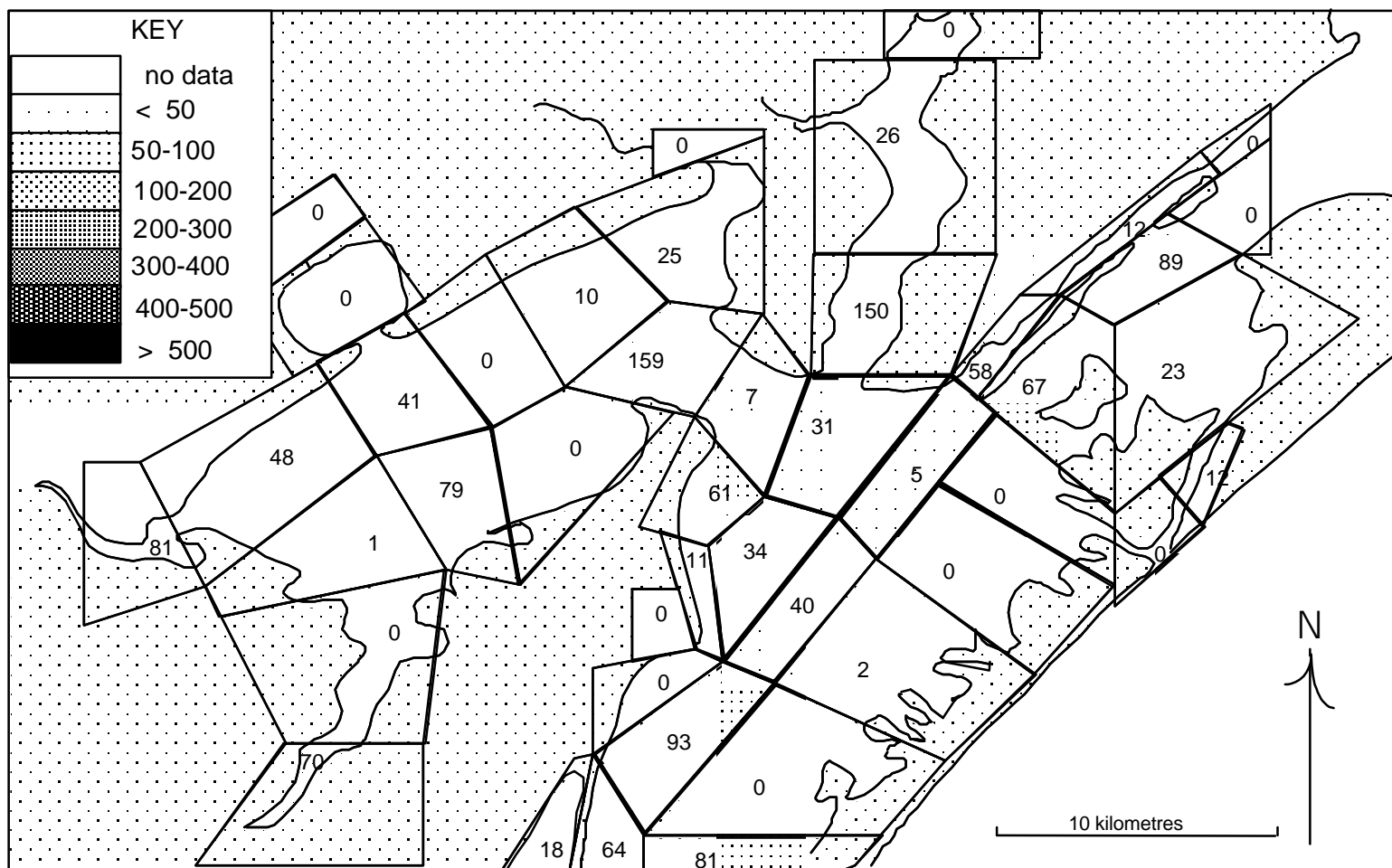


Figure 2-13. Sampling density of WQAMMN for Aransas-Copano system

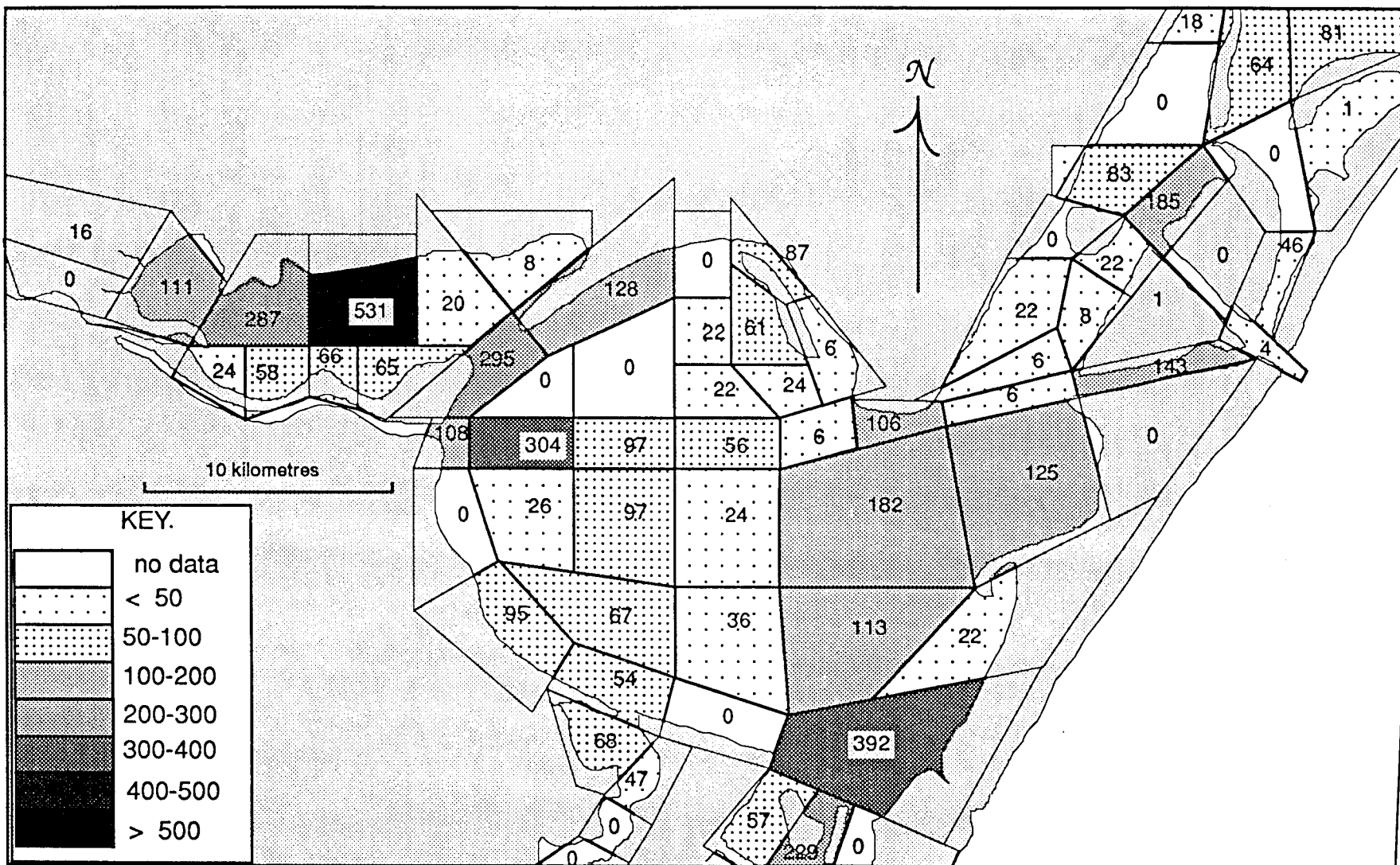


Figure 2-14. Sampling density of WQAMMN for Corpus Christi system

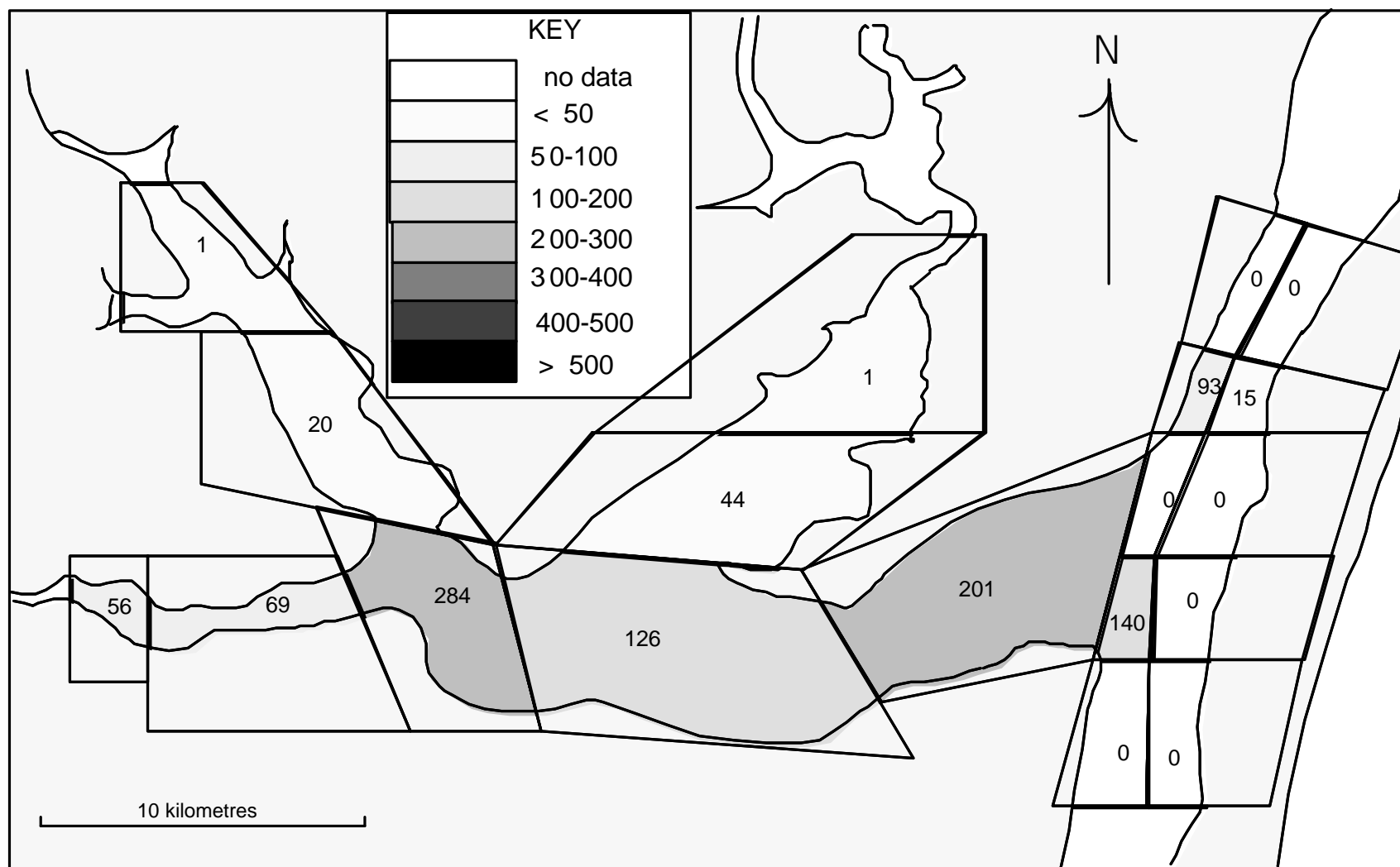


Figure 2-15. Sampling density of WQAMMN for Baffin Bay region



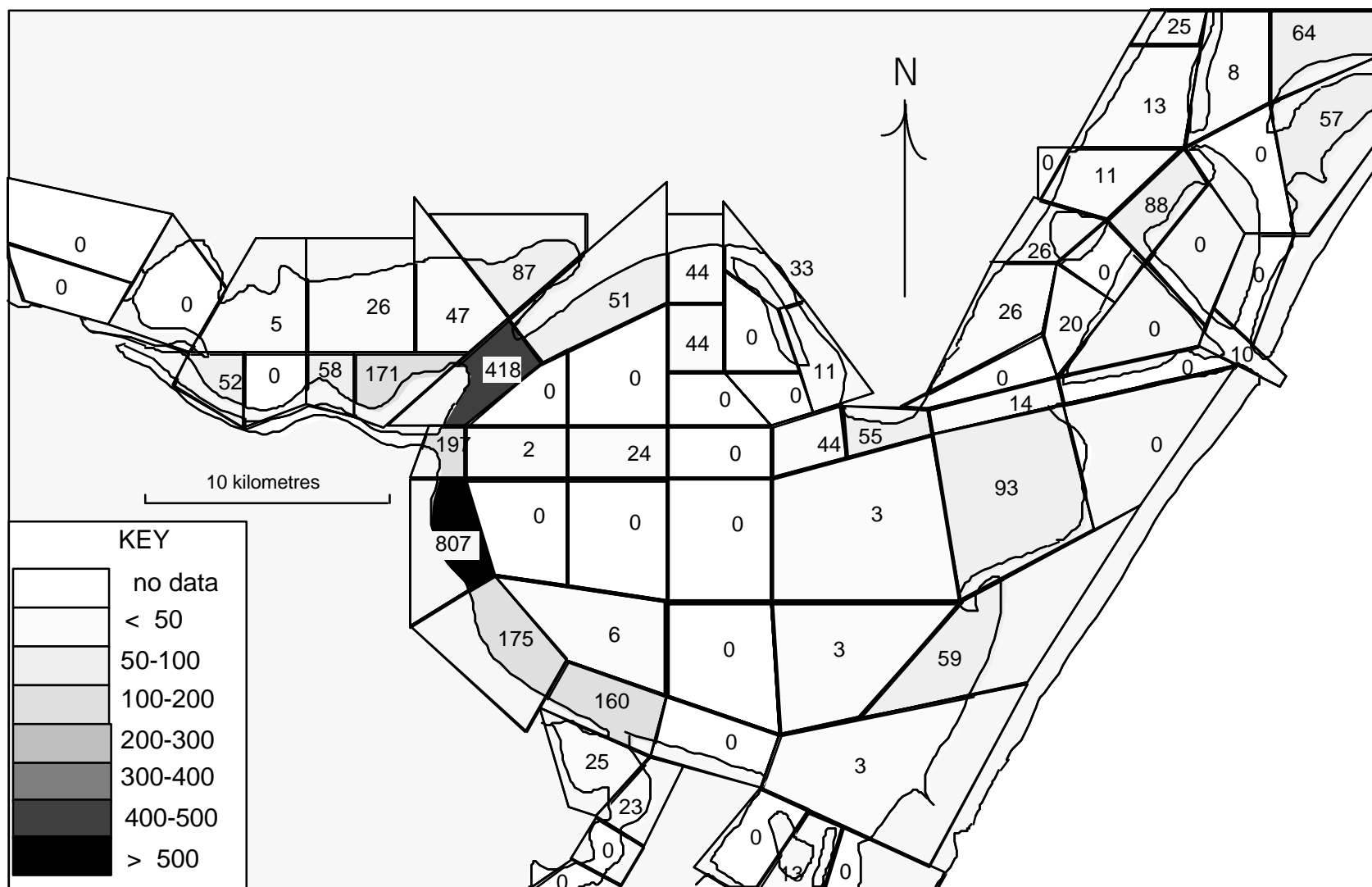


Figure 2-17. Sampling density of WQFCOLI for Corpus Christi system



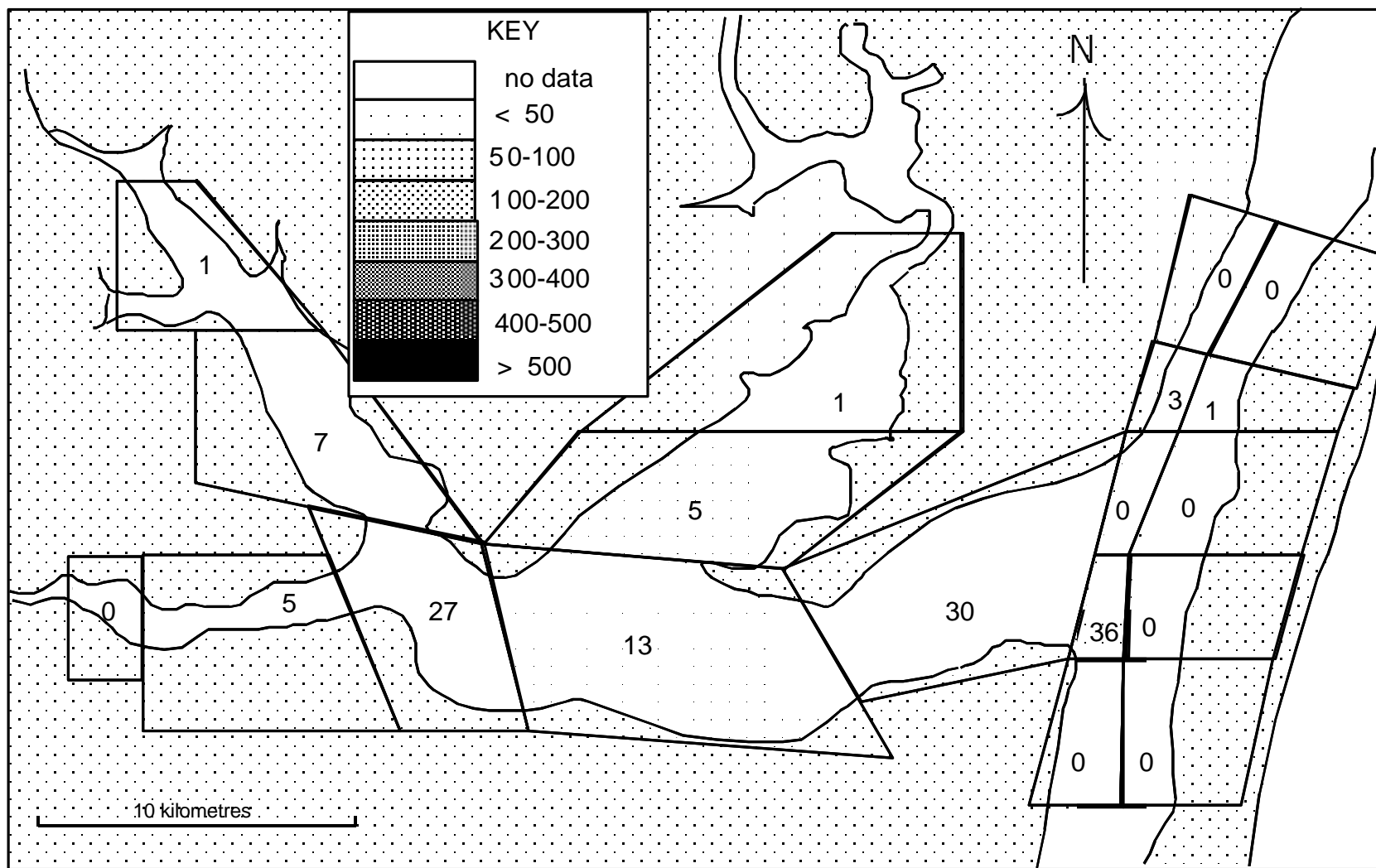


Figure 2-18. Sampling density of WQFCOLI for Baffin Bay region

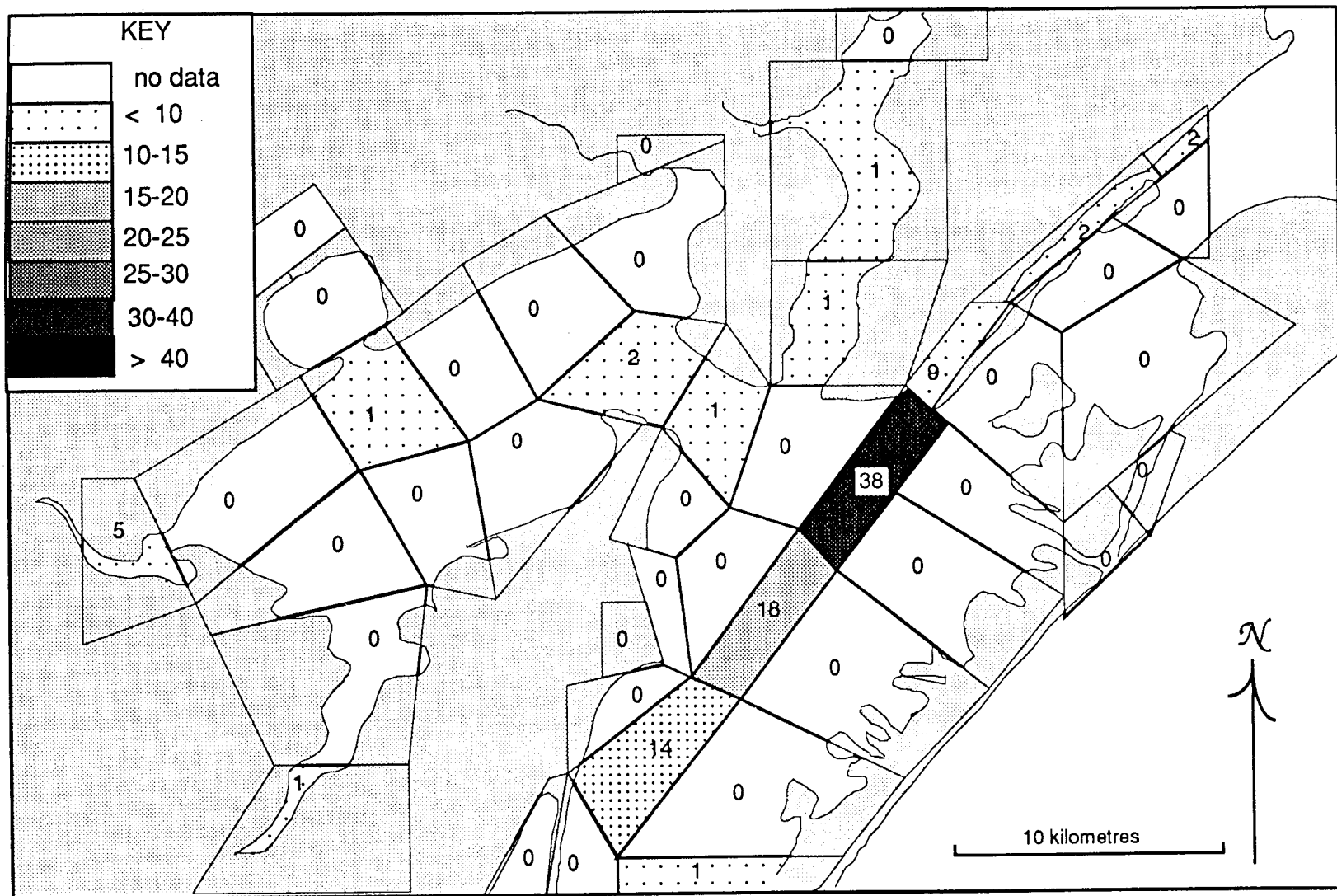


Figure 2-19. Sampling density of WQMETCDT for Aransas-Copano system



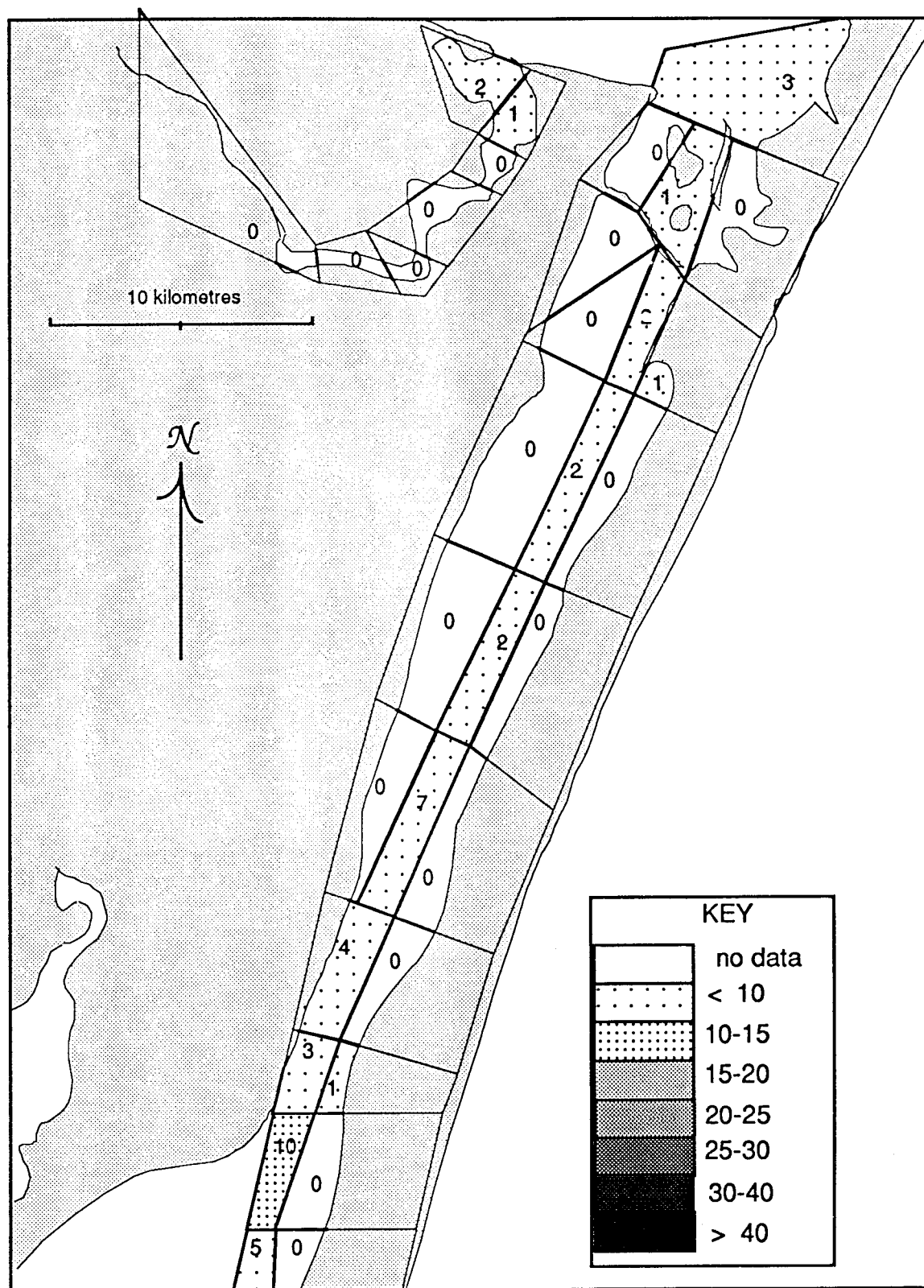


Figure 2-21. Sampling density of WQMETCDT for Upper Laguna Madre and Oso Bay

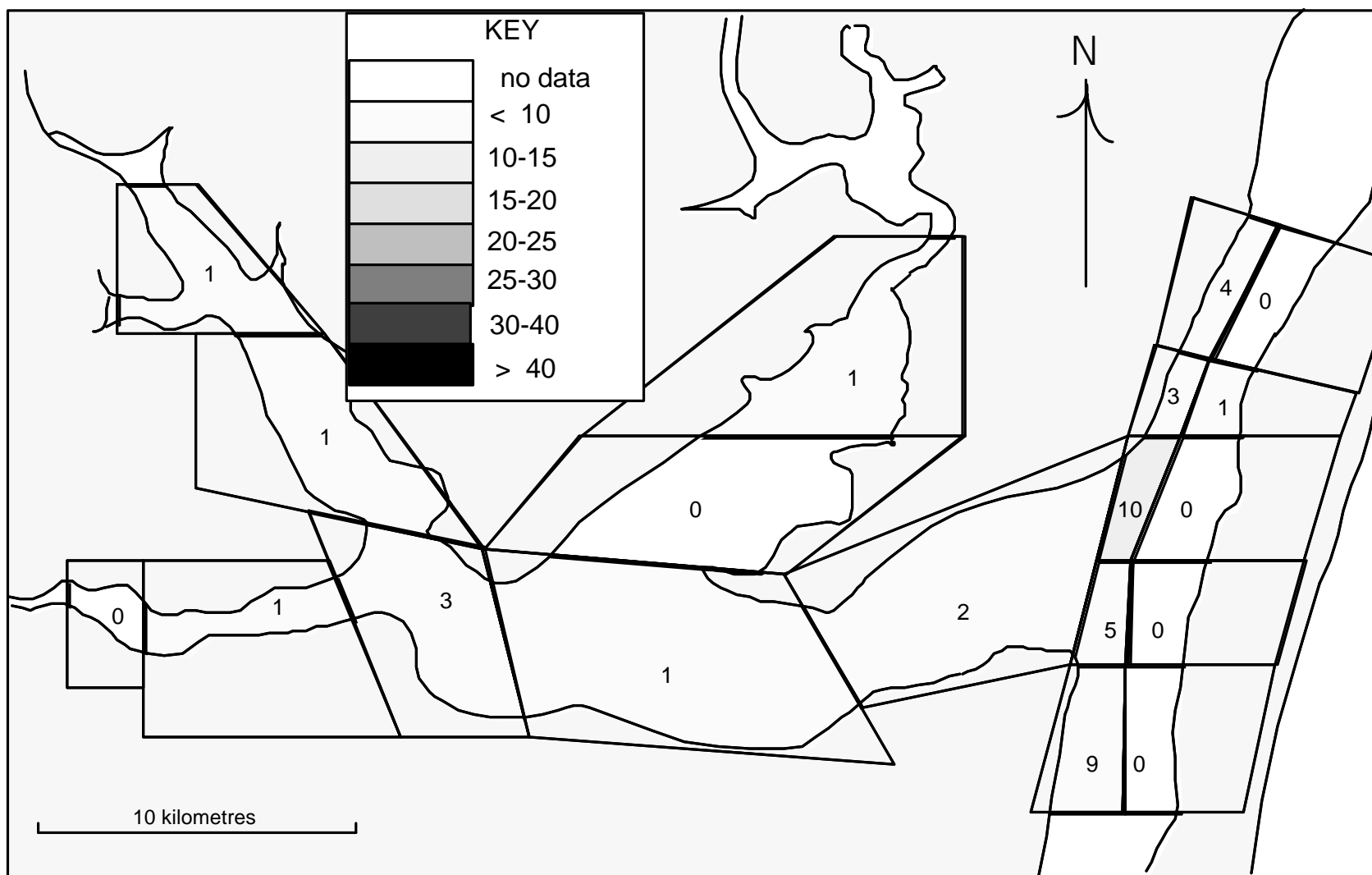


Figure 2-22. Sampling density of WQMETCDT for Baffin Bay region



Figure 2-23. Sampling density of SEDO&G for Aransas-Copano system





Figure 2-25. Sampling density of SEDMETZN for Aransas-Copano system



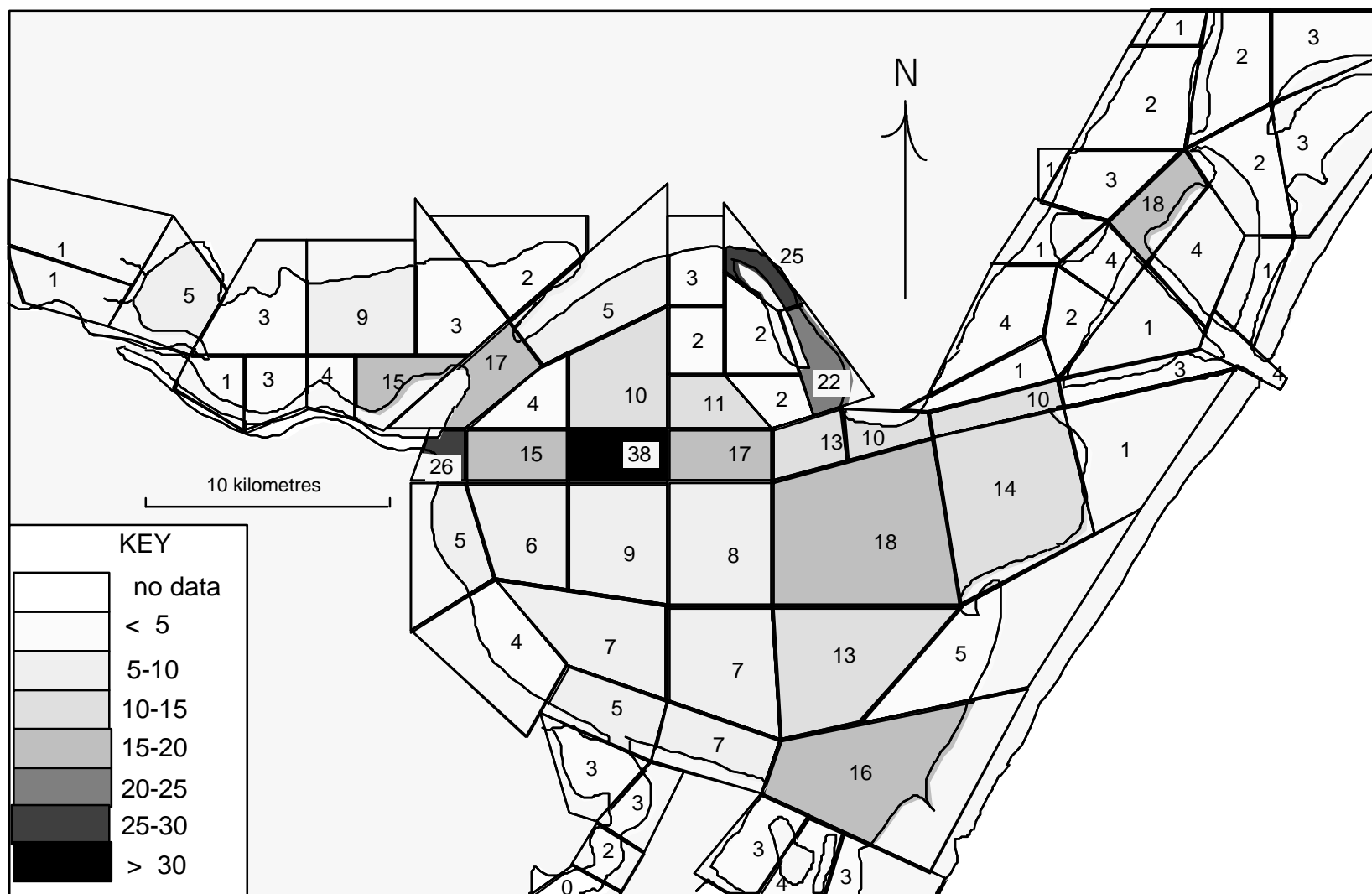


Figure 2-26. Sampling density of SEDMETZN for Corpus Christi system

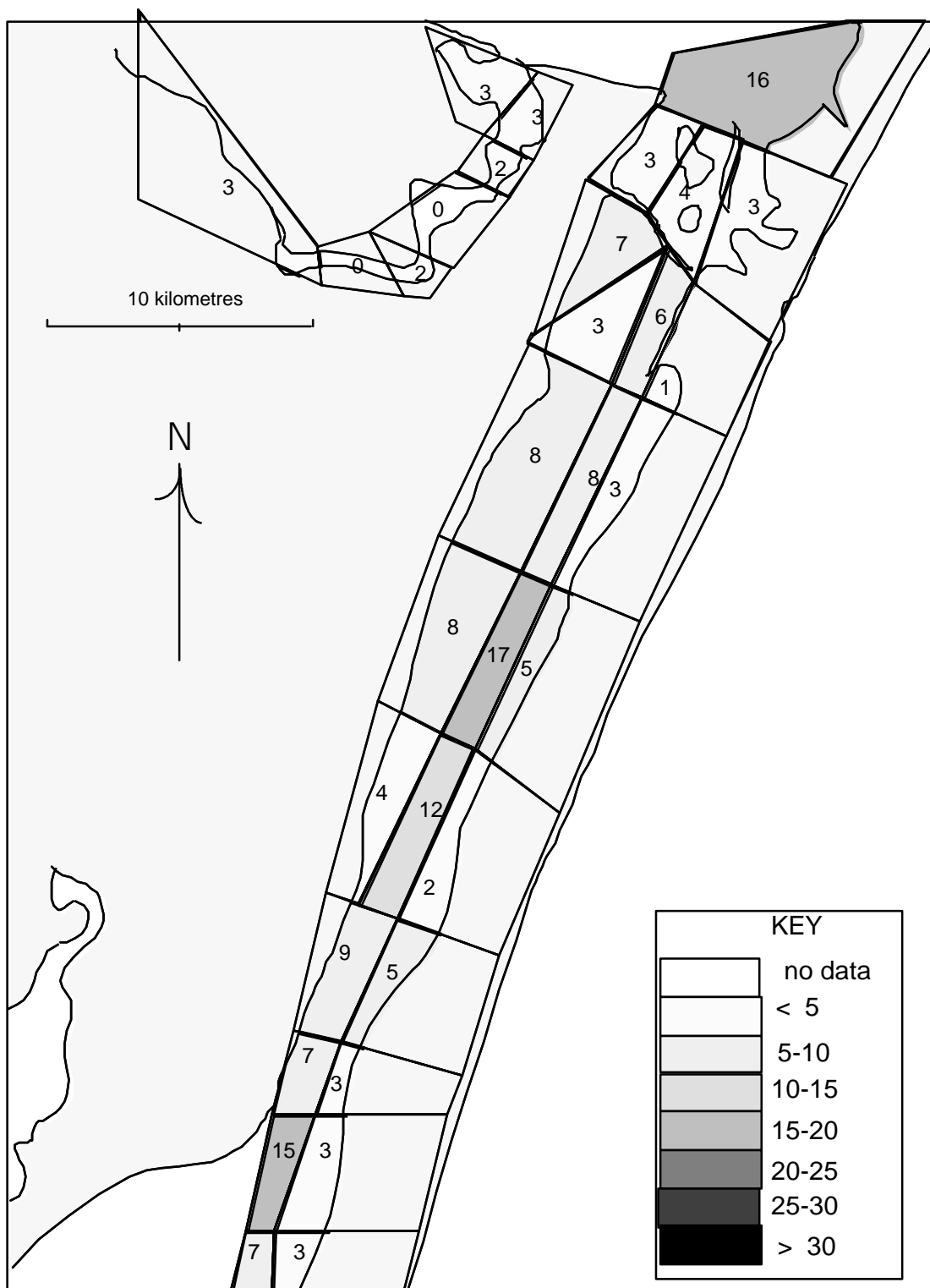


Figure 2-27. Sampling density of SEDMETZN for Upper Laguna Madre and Oso Bay

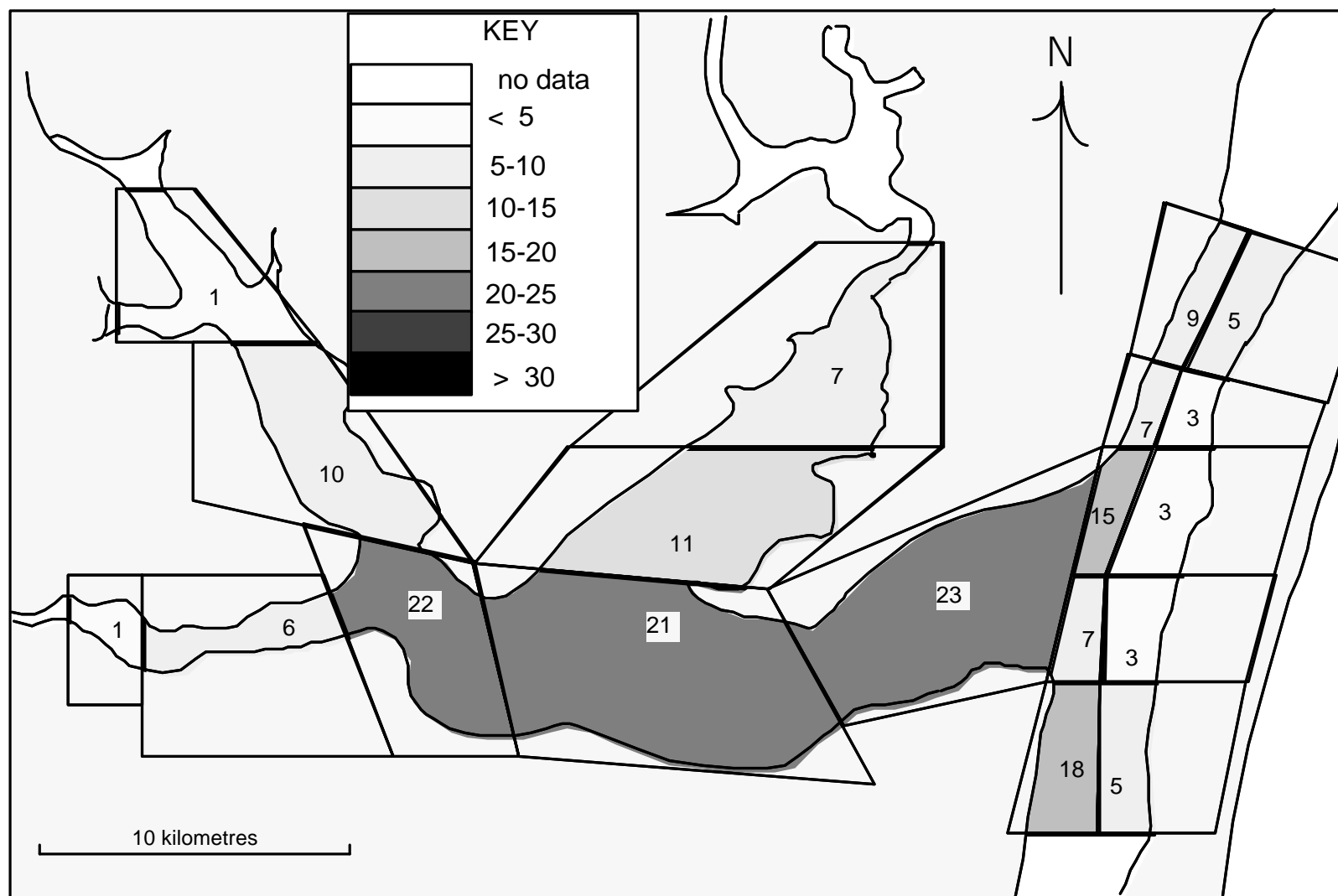


Figure 2-28. Sampling density of SEDMETZN for Baffin Bay region

space, some areas of the system 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 and for those areas with historical pollution problems. The period of record generally ranges over many years (see Tables 2-6 and 2-7) so the number of samples *per year* is a considerably smaller number.

The time history of sampling in the Corpus Christi Bay system since 1955 is roughly indicated in Figs. 2-29 through 2-36. 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. The intensity of coliform measurement in this data base generally increased since the 1960's, Fig. 2-31. While the coliform record encompasses nearly 40 years, it is really only half that good, since fecal coliforms have replaced total coliforms and there is no proxy relation between the two coliform indicators.

The increase in the frequency of trace constituent analysis since about 1980 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. 2-34, showing one metal, one insecticide, one PAH and total PCB's. Most of the parameters in Table 2-4 were introduced by this same technological innovation, because the GC/MS methodologies permit a large generation of parameters from a single sample/procedure, hence the coherence in number of samples since 1984. The best record is the extended DDT, obtained by combining reported "total" values with those proxied from the pp'-DDT isomer. 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).

With respect to tissue measurements, the available base of data can be summarized succinctly by the following:

- By far, the greatest quantity of data has been taken for the metals analytes.
- The most-sampled organism is the American oyster, followed by the blue crab, followed by speckled trout, red drum and black drum.
- Brown shrimp and white shrimp, relatively speaking, are rarely sampled. One sample of each appears in the data base.
- Among the organics, the greatest data base is that of the pesticides, especially the common commercial mixtures such as chlordane and toxaphene, and that of PCB's.
- Most of the organics in the suite of analytes have never been detected in the tissues of organisms.
- 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.

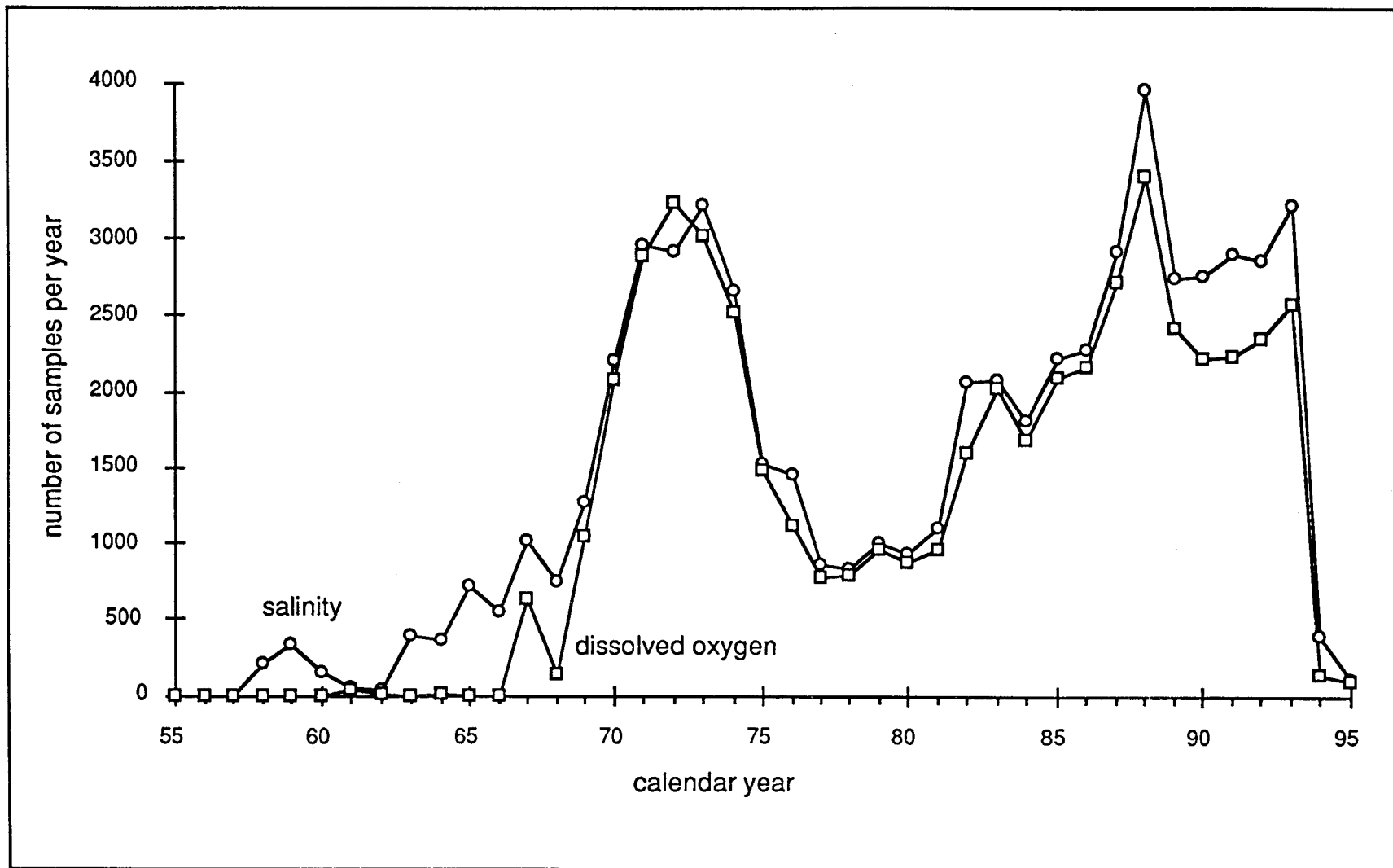


Figure 2-29. Data base annual measurements 1955-95, hydrographic parameters

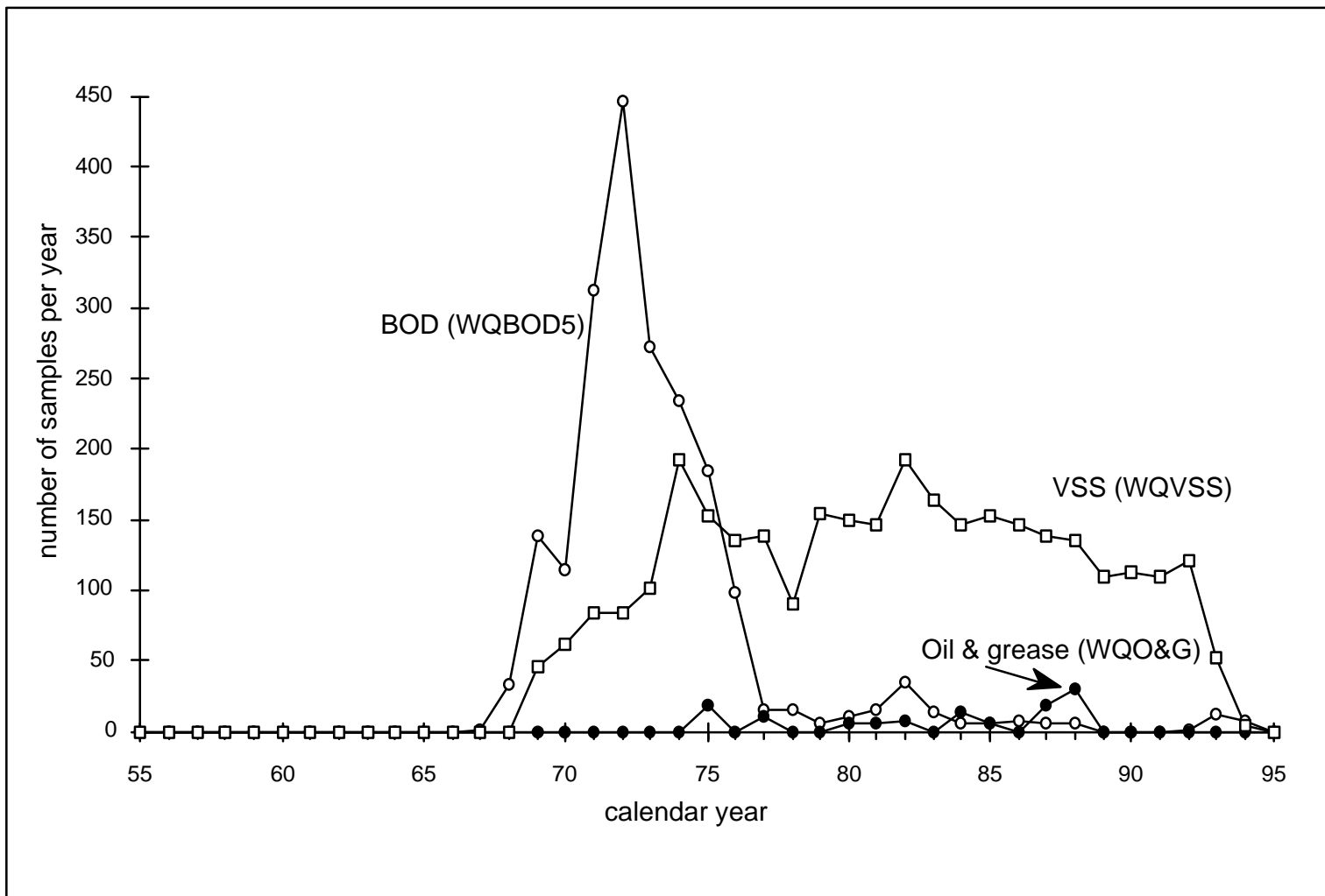


Figure 2-30. Data base annual measurements 1955-95, water-quality indicators

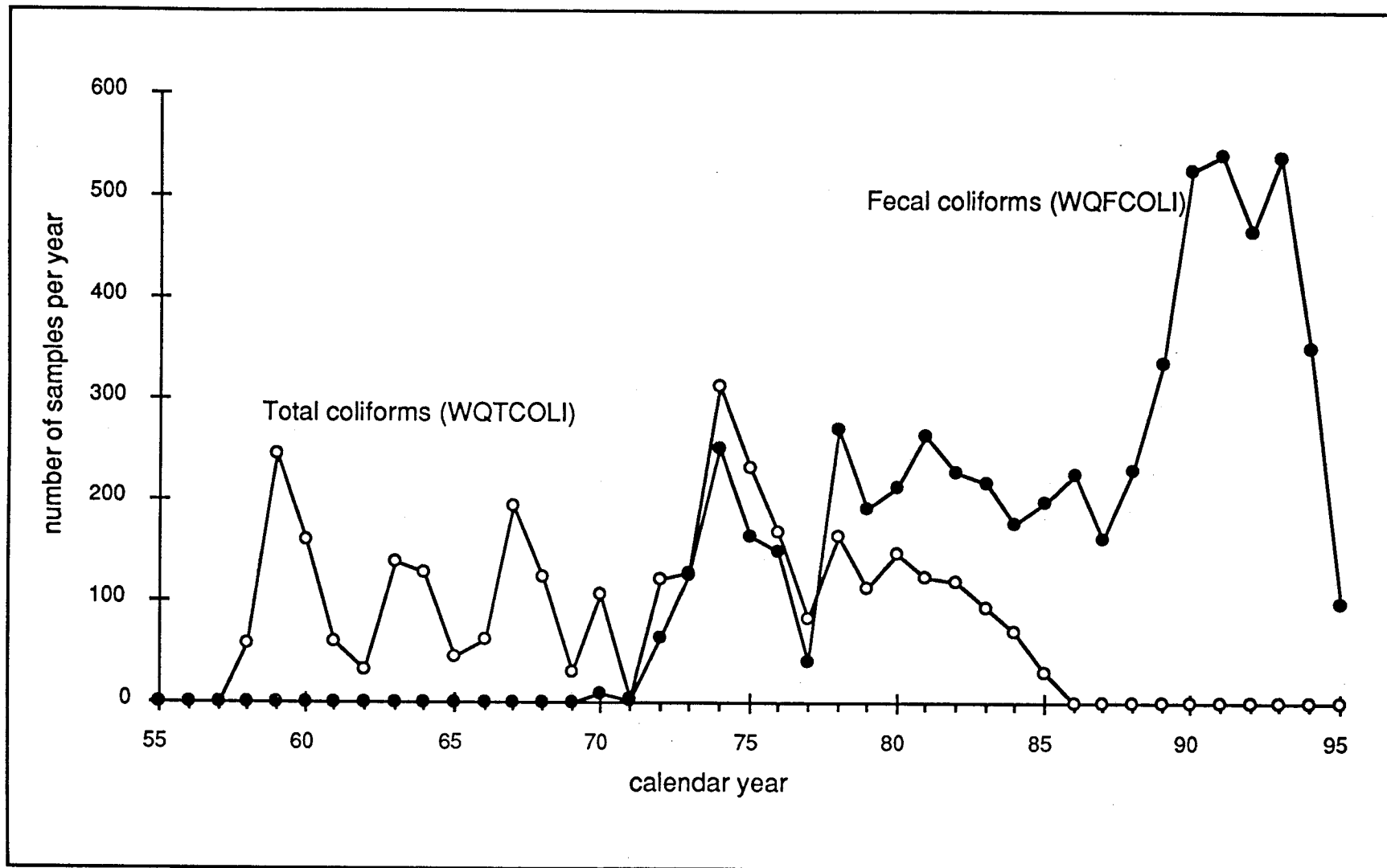


Figure 2-31. Data base annual measurements 1955-95, coliforms

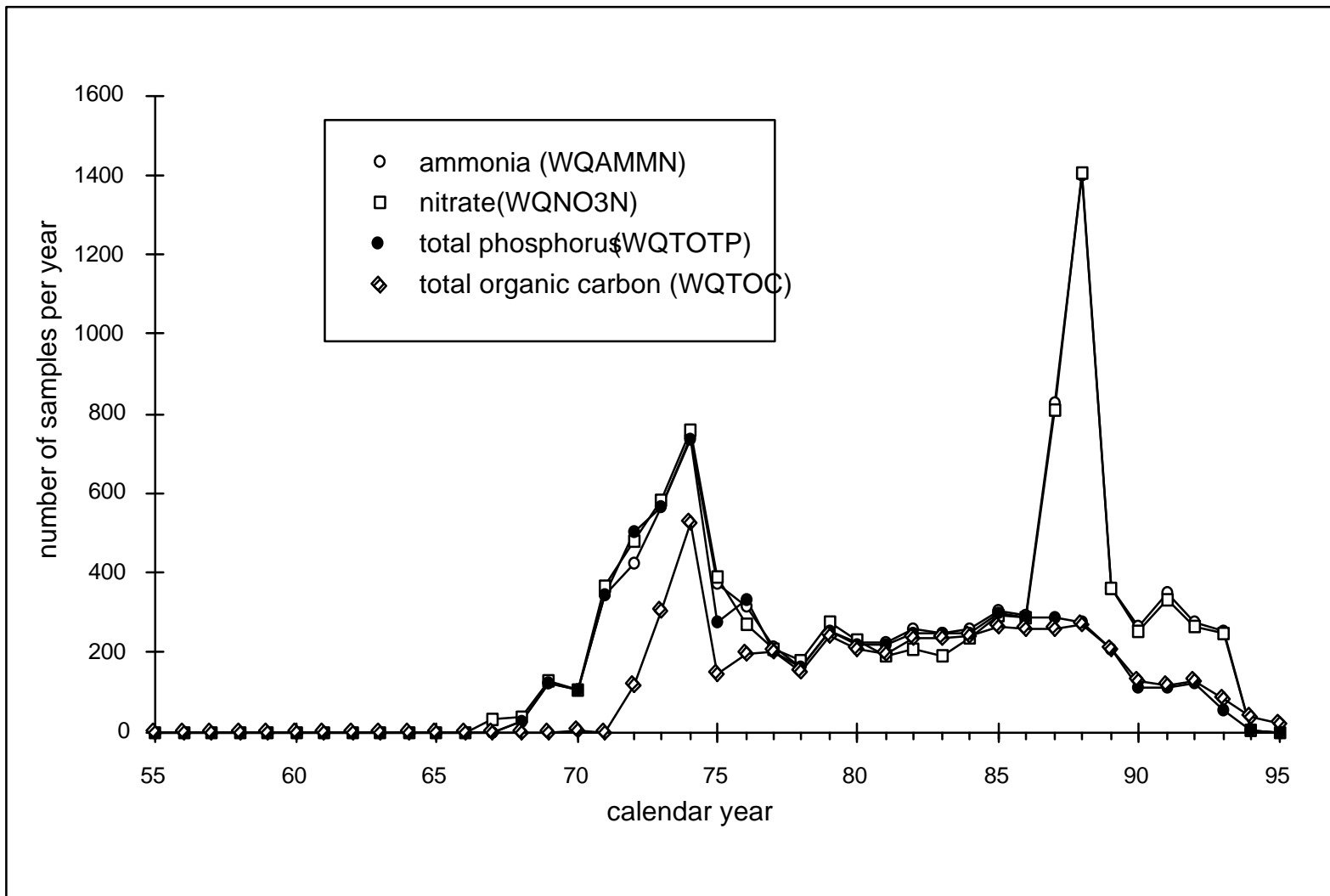


Figure 2-32. Data base annual measurements 1955-95, nutrients



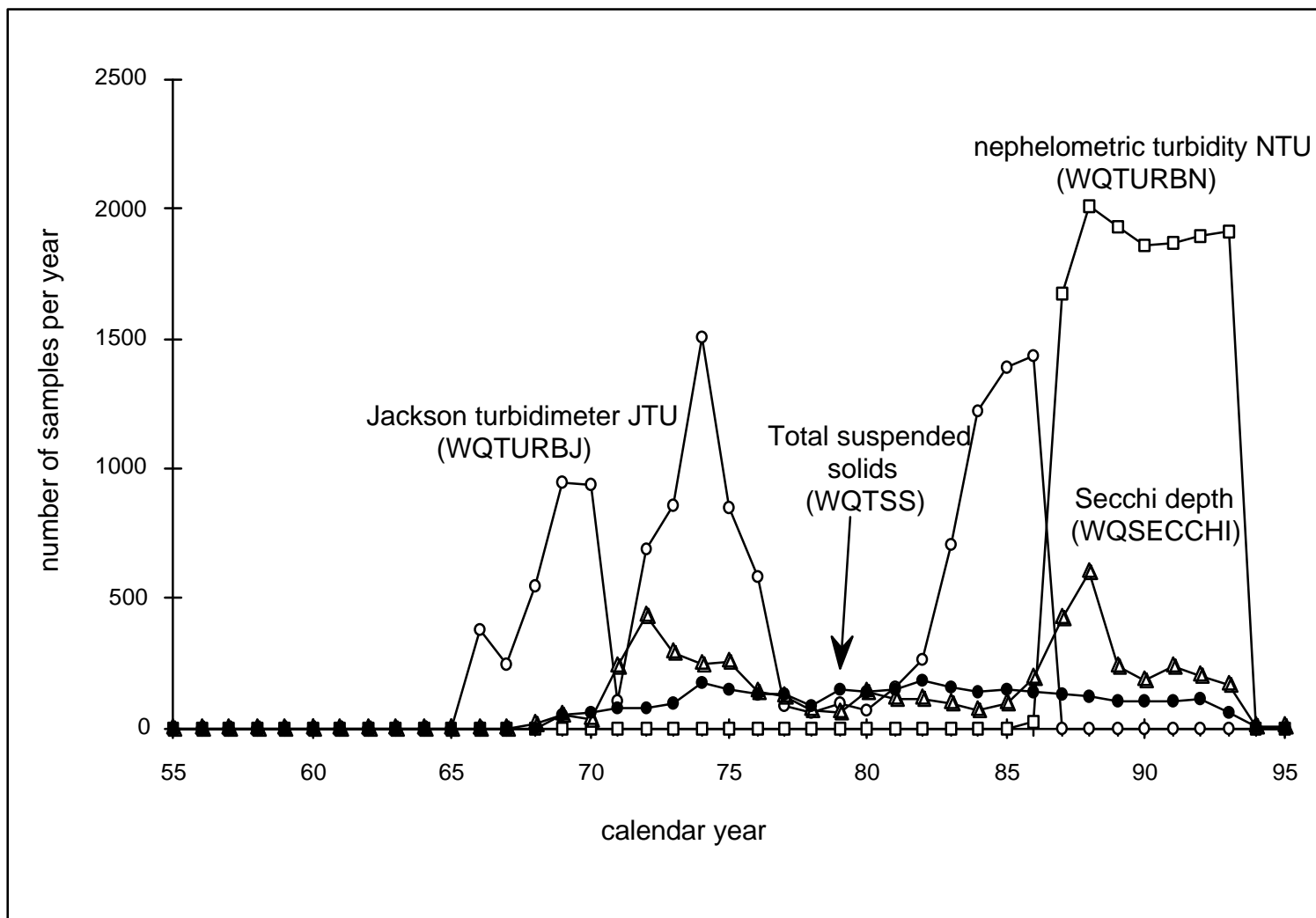


Figure 2-33. Data base annual measurements 1955-95, turbidity parameters

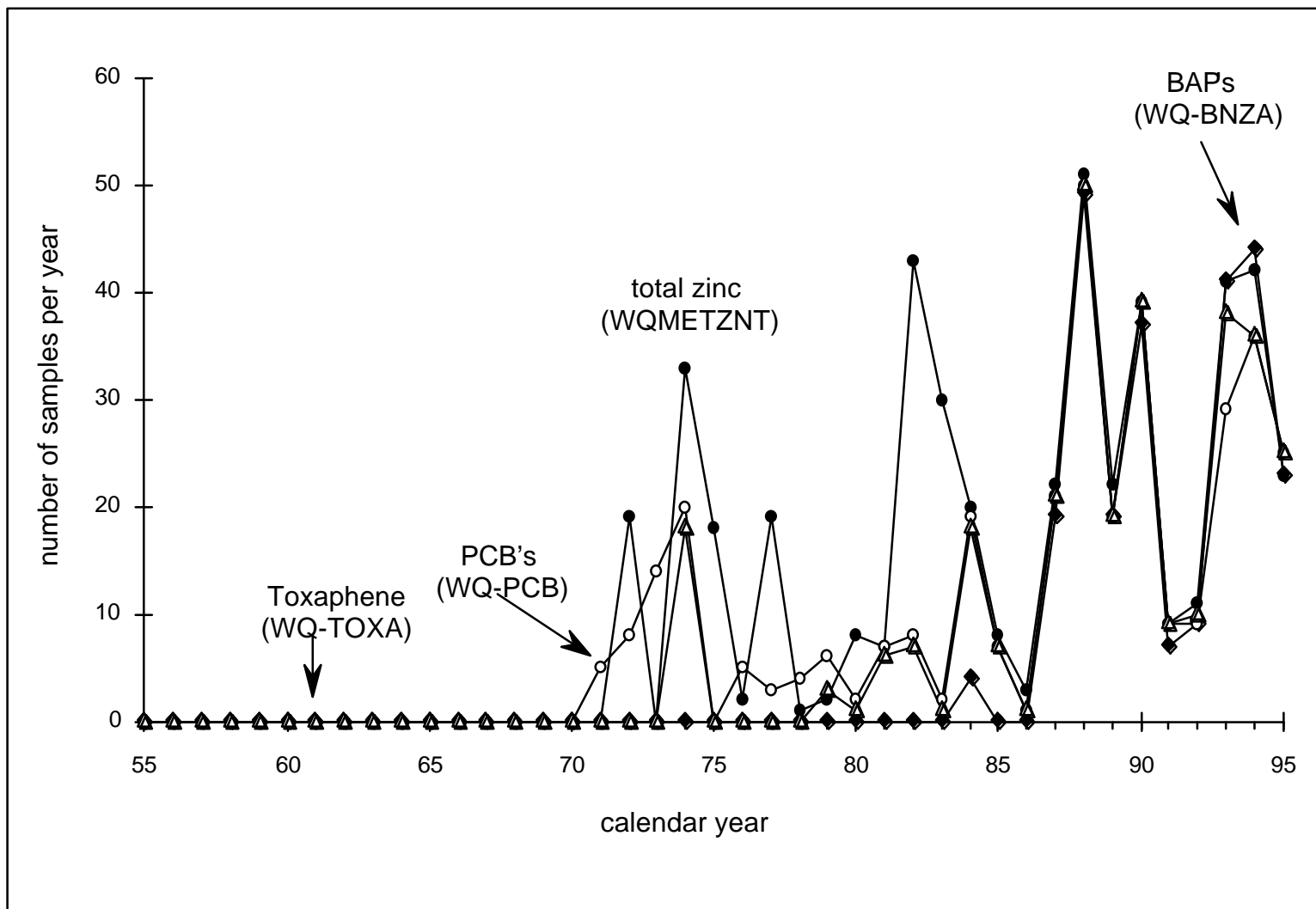


Figure 2-34. Data base annual measurements 1955-95, metals and toxics

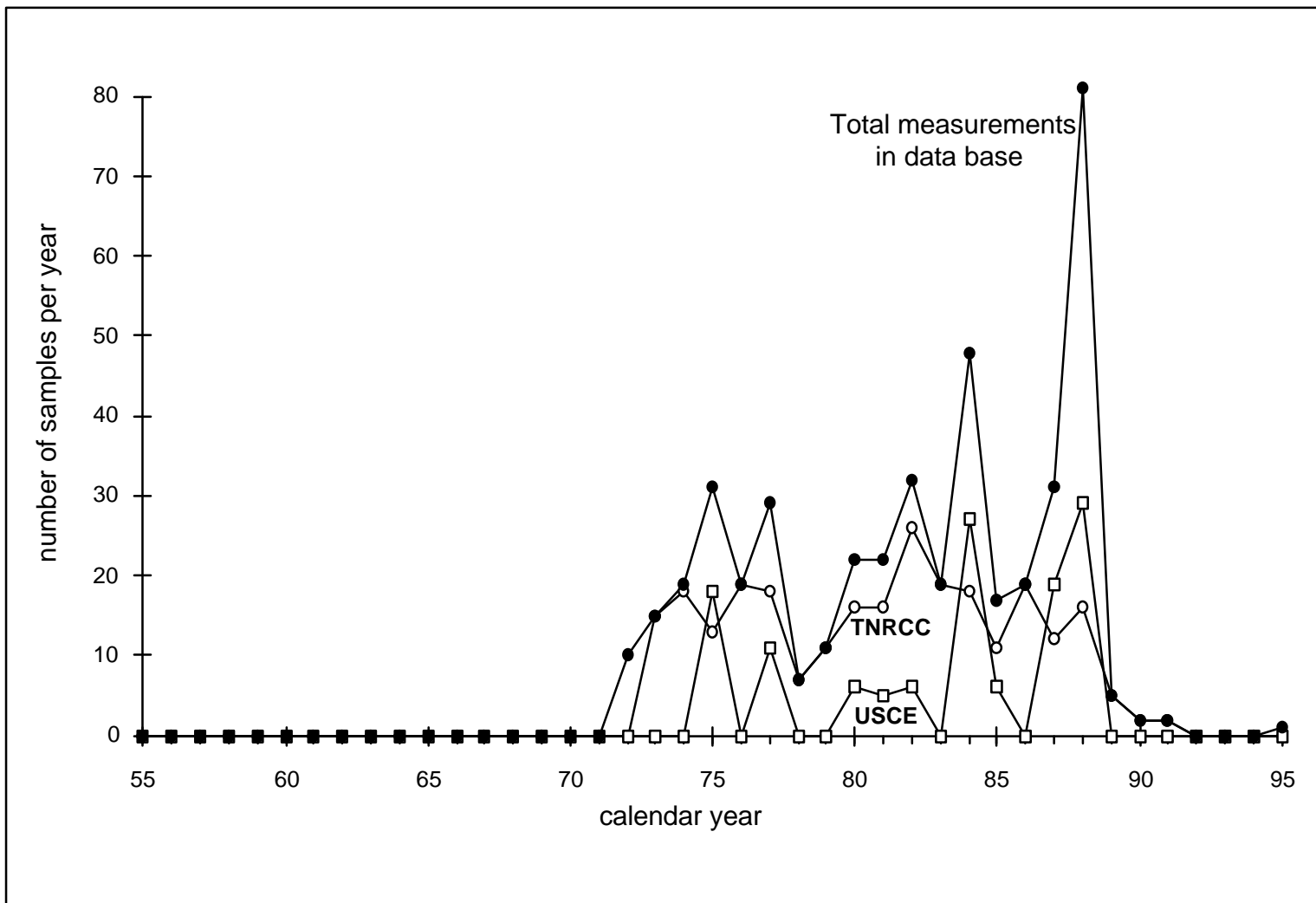


Figure 2-35. Data base annual measurements 1955-95, sediment oil & grease (SEDO&G)

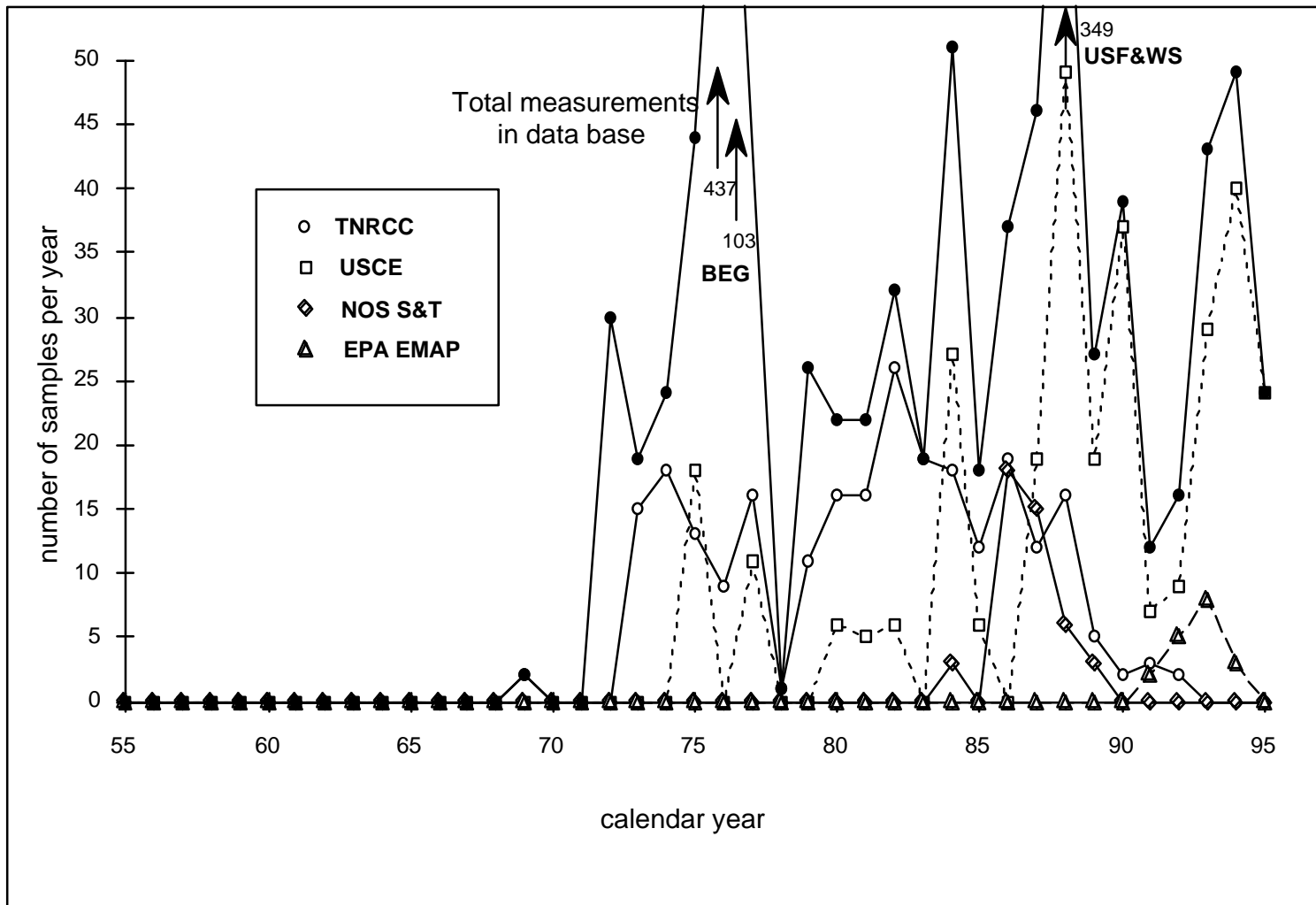


Figure 2-36. Data base annual measurements 1955-95, zinc in sediment (SEDMETZN)

One type of data conspicuous in its absence in this compilation is the data sets of salinity, temperature and dissolved oxygen obtained by the programs of robot or “sonde” data collection. The CCBNEP study area has been the primary region for research in this technology by two agencies, the Bays and Estuaries Program of TWDB and the Conrad Blucher Institute of Texas A&M Corpus Christi. These data files were provided to this project, and were given considerable study but were ultimately excluded from the present compilation. The reason is that the data are 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 over time due to these effects. Dissolved oxygen is even less reliable. While the research underway is addressing means to overcome these problems, and to properly calibrate the results, the fact is that in their present form, the data records evidence too much corruption to be incorporated into the present compilation. 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.

## **2.6 Deficiencies in data collection and management**

There are three fundamental dimensions to data collection: (1) range of parameters, (2) duration and intensity in time, and (3) distribution in space. The range of parameters determines what aspects of the Corpus Christi Bay environment are monitored, e.g. its hydrography, ecological and habitat quality, and potential limitations of human uses. Time continuity is of central importance in applying a data base to issues of environmental quality management, 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 system 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, and is much more important in improving statistical reliability than a reduction in the noise in the data. In Corpus Christi Bay, spatial gradients in concentration are created by circulation, intermixing of fresh water and seawater, and sources of pollutants. The network of data collection sites must be capable of resolving these spatial distributions in order to determine cause-and-effect, and to identify areas requiring special management actions.

To assess the adequacy of the data base in Corpus Christi Bay, each of these dimensions must be considered. They can be difficult to clearly separate, however. For convenience, in Section 2.6.1 we first address the parameters and their spatial distribution, with emphasis on the current practice of data collection, though of necessity this must include information on how this practice has changed in recent years. Then in Section 2.6.2 we assess present practices of data management, followed in Section 2.6.3 by consideration of management of historical data, which includes assessments of the adequacy of the historical data record.

### *2.6.1 Deficiencies in sampling strategy and data distribution*

The distribution in time of sampling of major classes of water and sediment parameters in the Corpus Christi Bay system was summarized in the previous section. The important conclusion is that since 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. However, there are some important exceptions to this statement, mainly in the collection frequency of hydrographic parameters and trace parameters. The recent increase in salinity and DO (temperature and pH being almost identical) shown in Fig. 2-29 has been mainly due to intensified data-collection of the Texas Parks and Wildlife Department Coastal Fisheries program. The importance of this data collection enterprise, in terms of the raw numbers of observations made, cannot be overstated. The increase in the frequency of trace constituent analysis, as noted above, was due to increased concern with metals and organic toxicants and to the advancement of analytical methodologies.

The decline of data on conventional contamination indicators since 1993 for water (Figs. 2-30 through 2-32) and since 1990 for sediment (Fig. 2-35) is due in part to the pipeline problem of sample processing at TNRCC, which has a backlog of many months (see Section 2.6.2 below). But it is also a very real decline, due to a reduction in the number of stations occupied, and the frequency of data collection, primarily of the TNRCC, but also of reduction of effort in several federal data collection programs implemented in the mid-to late-1980's. 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. 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.

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. For salinity, in 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 robot data logging and electrometric sensing now permit the recovery of nearly continuous, fine-scale time signals of these parameters. The work of the TWDB, the Marine Science Institute, and, more recently, Conrad Blucher Institute of Texas A&M University Corpus Christi represents major advances in sampling these fine time scales. This work is still very much in the research phase, and the data do not have a comparable reliability to those of traditional monitoring programs. But these technologies need to be brought to an

operational, routine state as soon as feasible. NB, such data acquisition does not replace routine sampling, since routine sampling provides far better spatial continuity than is economically achievable with robots, and is not subject to the vandalism and sensor degradation that 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. One program may emphasize COD and sulfides, say, while another examines phytoplankton and TOC, and a third may analyze BOD and chlorophyll-a. 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, which differs in turn from that of EPA's EMAP program. 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.

Inconsistency in measurements also occurs within the same data-collection program, but as a function of time. 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 naphthalene, acenaphthene and fluoranthene. (Further, in some programs, the specific PAH's vary from run to run.) A related problem is the discontinuation of various parameters. This is especially true for many conventional parameters such as oil & grease, BOD, volatile solids, and volatile suspended solids. Most of these can be determined inexpensively, and while they are rather imprecise, from the standpoint of establishing long-term trends, their continued monitoring does serve a purpose.

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 role particulates play in habitat quality, and in the sorption of nutrients and hydrophobic 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. Considering that a suspended solids determination is dirt-cheap (forgive the pun), one is mystified that it would be omitted as a companion measurement to a hyper-expensive GC-MS analysis for an extensive suite of trace parameters. Future sampling should include

routine measurement of suspended solids with every metals or hydrophobic organics sample. It would be even better to include a determination of grain-size distribution, as noted above.

In addition to their relation to suspended solids, interpretation of organic-contaminant data frequently 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. Again, we see the lack of a relatively inexpensive conventional measurement (TOC) seriously limiting the utility of a much more expensive GC-MS determination.

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 significant 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 2-7 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). 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 anomalous zero values).

The most important monitoring project that routinely collects sediment quality data throughout the system is that of the TNRCC. USCE collects data at irregular intervals near its navigation projects. Both the NOS Status & Trends Project 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. The BEG Submerged Lands Project and the EMAP programs are the only instances in the Corpus Christi Bay data record in which sediment texture and metals data were both obtained. Only the BEG collected a respectable spatial density of samples, but even in this program only a small minority of the samples were *paired*, i.e. chemical and textural analysis run on aliquots of the same sample. For the other sediment data, there is no textural analysis. Once more the lack of a companion parameter, an inexpensive measurement compared to instrumental chemistry, seriously limits that value of the analytical result: this situation is inexplicable.



Much of the historical data for metals and trace organics in water and sediment has been corrupted by inattention to detection limits. In the historical data record, detection limits are frequently reported in error (perhaps not determined as a part of the analysis) or, worse, zero values of concentration are reported. While the present situation is much improved, due to the enforcement of uniform laboratory analytical procedures, we found several instances in this data compilation of questionable detection-limit determinations in recent projects (Ward and Armstrong, 1997b). But more problematic is the universal practice of censoring the data by replacing the analytical determination with the statement that it is "below detection limits." Considering the investment in time and expense to obtain a sample and subject that sample to precise and complex instrumental analysis, it is inexplicable why the analytical result would be discarded because the risk of a Type-I error exceeds 1%. Much useful information that could have been incorporated into a statistically sound analysis has been lost by the practice of censoring. It has also engendered a cottage industry of methods that attempt to recover that information by statistical hocus-pocus.

With respect to the spatial dimension of sampling, of course the optimum sampling distribution would be a dense, uniform network throughout the system. This is rarely possible, especially for a monitoring program. 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. First, stations are placed where the activities within the purview of that agency are most likely to be encountered. 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 primarily subject to shellfish harvesting. The Corps of Engineers concentrates its monitoring activities in the proximity of ship channels. All of these may be useful locations. The problem is that the remaining areas of the system are not sampled.

Second, the precise location of the stations may be dictated by ease of location, or maintenance of tradition, rather than its maximal utility. 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 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 should ideally 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 one 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.

The sampling density figures given in the preceding section evidence a high degree of heterogeneity in sampling. As might be expected, the Corpus Christi Bay component has the richest data base, and the upper and lower systems of Aransas-Copano and Baffin are sampled much less frequently. The sediment samples are particularly poorly distributed, with the outer bays not being sampled for many parameters in over five years. 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 water and sediment quality as a basis for habitat assessments, the distribution is not appropriate.

The spatial distribution of sediment 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 2-1). 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 in the late 1980's (Project 27 in Table 2-1), 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.

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.

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 of the technical report (Ward and Armstrong, 1997a). 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 to this project. 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.* radio-navigation systems and 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. In this compilation, in all of those surveys using these methods, we found substantial errors in position (i.e., tens of kilometres). Fortunately, with some effort and the assistance of the agency or investigator supplying the data, they could be fixed (sometimes by fallback to “eye-balled” locations). In the present context, while we believe this to be welcome technology, we note that, the vaunting of vendors notwithstanding, there are sources of error. These are compounded by a proclivity to utilize these systems without adequate understanding of their principles of operation and their limitations, to regard the results with unquestioning veneration, and to employ no verifying independent location data.

#### *2.6.2 Deficiencies in management of modern data*

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. 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 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 much of 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, both of modern data bases, i.e. more-or-less contemporary data streams, and of historical data bases.

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 Galveston Bay, most of them apply as well to the Texas coast in general and to Corpus Christi Bay in particular. Also, though 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 brief discussion and examples drawn from the present study. Reference is made to the technical report (Ward and Armstrong, 1997a) for additional detail.

### ***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, 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 2.1.1 above) 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 and commendable 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. 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 sort of problem is not a lone example, but a constant frustration in the preservation of digital data bases.

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

In Section 2.1.1 was noted 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. 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, 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 the failures to flag BOGAS data and to check the calculation.

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 the pre-existing data, 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 spreadsheets (but in many cases only hard-copy printouts have survived). More recent data are maintained in data base manager software. 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 proliferation of software products, the promotional device of American software purveyors of issuing updates and

new releases at a diarrhoetic rate, and the gullibility of computer managers in succumbing to these promotions, 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 ferreted out 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 CCBNEP study 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.

### *2.6.3 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 contribute to this loss of historical data. 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. Even extant data management programs are precarious, in that they are sensitive to shifts in organizational emphasis. An office purge is forever.

The prominent example of this in the Corpus Christi Bay data record is the loss of the Phase III data reports from the USF&WS study of Matagorda Bay in the late 1970’s. The Information Transfer Office of USF&WS, whose sole purpose is archiving and dissemination of USF&WS reports, was unable to produce these documents.

#### ***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 Department of Health is to determine the suitability of fish and shellfish for human consumption. The present condition is always primary. The Texas Natural Resource Conservation Commission is 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 types of agencies that are largely responsible for the bulk of data collection within the Corpus Christi Bay system.

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, open-bay surveys, parameters unknown, 1960?-1965
- Texas Game, Fish & Oyster Commission surveys, 1950?-1970

The loss of the early City-County Health Department and the Corpus Christi Sewer Department surveys is particularly frustrating, as these would have extended the record back in time substantially, especially for conventional water-quality variables. 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 form of the new administrative function of Records Management, in current fashion in the state agency headquarters. These “records managers,” who are typically nontechnical and seem to 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 world) 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 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 one USGS staff person in Denver, 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.

#### ***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 mag-tape 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, the sliderule, and the Magcard, the technological hardware support is no longer available.

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. A 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.)

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 permanent media begin to appear, e.g. the CD-ROM, EM loss is eliminated but the possibility of simple physical loss becomes greater. The key to minimizing this hazard is to create multiple copies.

## ***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.

This page intentionally left blank.