Database documentation: kina

K. A. Mackay & D. O. Fisher

NIWA Fisheries Data Management
Database Documentation Series

Revised January 2016

Revision History

Version	Changed by	Reason	Date
1.0	K Mackay, D Fisher	Initial release	1993
1.1		Revised data structures section and renumbered headings to standard layout.	Nov 2000 (or prior)
1.2	D Fisher	MPI version	Jan 2016

Note this revision history section was added in January 2016 so it may not include all revisions prior to then.

Contents

1 Introduction to the Database Document series 4
2 Kina Population Surveys 4
2.1 Data sources 4
2.1.1 Commercial Catch and Effort 5
2.1.2 SOP Catch Sampling 5
2.1.3 Marine Research Dive Surveys 6
2.2 Data Validation 6
3 Data Structures 8
3.1 Table relationships 8
3.2 Database design 9
3.2.1 Commercial catch and effort data structures 9
3.2.2 Biological data structures 10
3.2.3 Grid references 11
4 Table Summaries 12
5 kina Tables 13
5.1 Table 1: t_catch_site 13
5.2 Table 2: t_sample 14
5.3 Table 3: t_test_dia 15
5.4 Table 4: t_biological 16
5.5 Table 5: t_grid_ref 17
6 References 18

List of Figures

Figure 1: Entity Relationship Diagram (ERD) of the kina database. 7

1 Introduction to the Database Document series

The National Institute of Water and Atmospheric Research (NIWA) currently carries out the role of Data Manager and Custodian for the fisheries research data owned by the Ministry for Primary Industries (MPI) formerly the Ministry of Fisheries.

This MPI data set, incorporates historic research data, data collected by MAF Fisheries prior to the split in 1995 of Policy to the Ministry of Fisheries and research to NIWA, and data collected by NIWA and other agencies for the Ministry of Fisheries and subsequently for MPI.

This document is a brief introduction to the kina catch sampling database **kina**, and is part of the database documentation series produced by NIWA. It supersedes the previous documentation by Mackay & Fisher (1993)[1] on this database.

All documents in this series include an introduction to the database design, a description of the main data structures accompanied by an Entity Relationship Diagram (ERD), and a listing of all the main tables. The ERD graphically shows how all the tables link together and their relationship with other databases.

This document is intended as a guide for users and administrators of the **kina** database.
This database has been implemented as a schema within the Postgres database called **fish**.

Access to this database is restricted to specific nominated personnel as specified in the current Data Management contract between the Ministry for Primary Industries and NIWA. Any requests for data should in the first instance be directed to the Ministry.

2 Kina Population Surveys

2.1 Data sources

This database results from the experimental fishing of kina (*Evechinus chloroticus*) off Fiordland.

Previous surveys of kina in Dusky Sound revealed abundant stocks, but insufficient information to predict sustainable yields from any fishery (McShane & Naylor 1991). To gain a detailed stock assessment of kina, a small number of fishers (participants) have been granted permits to harvest 1000 tonnes per annum of kina between them, commencing in the 1992-93 fishing year in Dusky Sound.

Fishing is restricted to free dive harvesting within the fishdown area, with permanent non-fish (control areas) established within the Sound. Each participant can have several divers fishing for them. Small-scale records of catch and effort are collected from the participants, which can be verified by scientific observers. Observers also record morphometric data from random samples of the catch.

[1] MACKAY, K.A. and FISHER, D.O., 1993: Marine Research database documentation 7: kina. *MAF Fisheries Internal Report No. 213.* 19p.

During the survey, Marine Research divers from the Kina Research Group regularly survey sites within the fishdown and control areas.

Stratified random surveys of kina in Dusky Sound (McShane *et al*. 1993) were conducted prior to the commencement of fishing. These surveys allow changes in kina stock structure and disturbances to the marine community caused by fishing to be accurately assessed.

As at September 1993, these surveys may be repeated at other sites within New Zealand, such as the Marlborough Sounds.

Data therefore come from three sources: (a) commercial catch and effort, (b) the Scientific Observer Programme (SOP) catch sampling, and (c) research dive surveys.

2.1.1 Commercial Catch and Effort

Each participant in the Kina Development Programme must complete daily catch log sheets at the end of each fishing day. This log is over and above the catch and effort landing returns that must also be completed. On this daily log sheet the participant records the landed catch weight, total time spent diving, and position for each catch site used that day by a diver; i.e., there is one daily log sheet for each diver. Hence, differences in catch and effort between individual divers can be estimated. Catch and effort for kina landed from a single dive site, but from multiple divers, are divided equally between the divers concerned.

2.1.2 SOP Catch Sampling

Scientific observers present on fishing boats during the Kina Development Programme collect biological data from the daily catches.

Each observer takes representative samples from a participant's daily catch. Samples are taken only from catches that can be traced back to a single dive site, and preferably from a single diver. Samples from a single dive site and diver can be linked back to the catch details as listed in the daily log sheet. Samples from a single dive site but multiple divers cannot be linked back to any catch details, but are still very useful in determining stock structure.

From a sample, an observer collects two sets of information:

1. The observer collects length frequencies of the commercial catch sample. This is achieved by measuring the test (shell) diameter (to the nearest millimetre) of about 200 individual kina. Each length frequency sample contains header information including, the participant's name, fishing date, grid reference, diver's names (there may be more than one diver), etc.
2. The observers are also required to provide information on kina morphometrics and biology (length/total weight/roe volume). Observers are instructed to take a random sample of 25 individual kina from the original sample of 200. For each kina, observers measure the test diameter to the nearest millimetre, weigh the whole animal to the nearest gram, record roe volume to the nearest millilitre, and record the colour, hence quality, of roe on a scale of 1 to 5 (with 5 representing top quality).

Observers also complete a daily catch log sheet, which is very similar to the participant's log sheets. Therefore, on most fishing days there are two sets of catch information. The compulsory filling of the logs by the participants ensures catch data for every fishing day, and the observer filled catch logs allow verification of the participant's logs. The **kina** database contains mostly observer collected catch data. Participant catch data are substituted only when there is no observer coverage.

2.1.3 Marine Research Dive Surveys

At randomly selected sites within a defined area, currently within Dusky Sound, two divers each descend to a randomly chosen depth (0-10 metres). Beginning in a randomly chosen direction, each diver turns a 1 metre square quadrat over 25 times. Each quadrat is thoroughly searched and all kina are counted and collected. The collected kina form the length frequency sample for that site. For some sites, a sub sample of 20 kina are randomly selected from the mature sized fish in the sample for a morphometric and biological examination. Data are collected and recorded as for the SOP sampling.

2.2 Data Validation

While the **kina** database structure enforces data validation and integrity with the use of referential constraints and range checks, the data also goes through a rigorous data validation and error checking process before being inserted.

The punched data are reformated using the **checkq**[2] program, which performs range checks, and the data are checked for invalid entries. The data are then loaded into temporary tables in the database where cross checks are made including checking that the grid references are valid, i.e., not on land or in too deep water, nor outside the valid area. Strata are allocated based on the grid references.

Each SOP sample is also checked to ensure it has an accompanying catch record, and each biological record has a corresponding length frequency record.

[2] See the local Unix manual page on **checkq**.

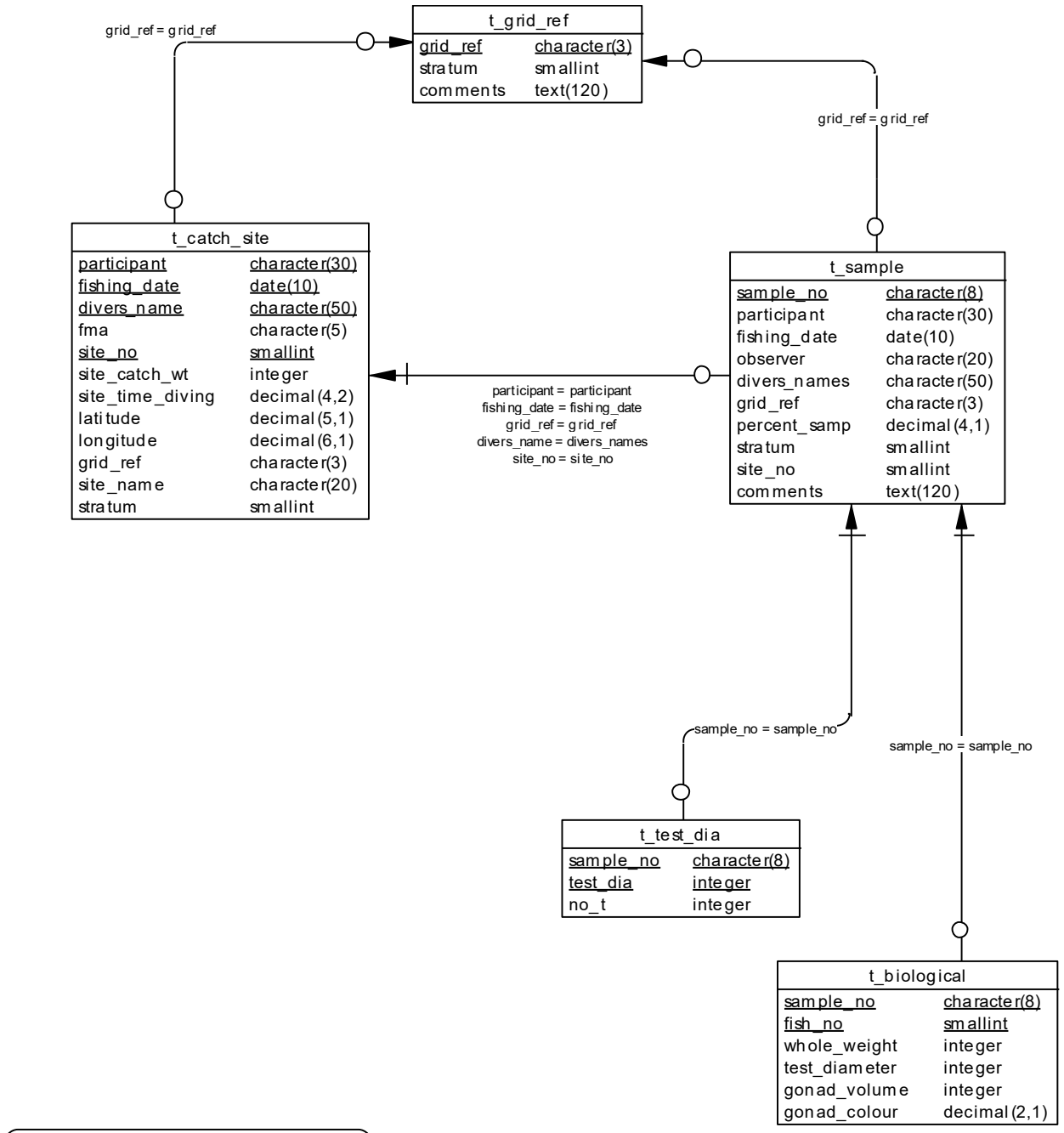

Figure 1: Entity Relationship Diagram (ERD) of the kina database.

3 Data Structures

3.1 Table relationships

This database contains several tables. The ERD for **kina** (Figure 1) shows the logical structure[3] of the database and it's entities (each entity is implemented as a database *table)* and relationships between these tables and tables in other databases. This schema is valid regardless of the database system chosen, and it can remain correct even if the Database Management System (DBMS) is changed. Each table represents an object, event, or concept in the real world that is selected to be represented in the database. Each *attribute* of a table is a defining property or quality of the table. All of the table's attributes are shown in the ERD. The underlined attributes represent the table's primary key[4].

Note that Figure 1 shows the main tables only. Note that most tables contain foreign keys[5]. These foreign keys define the relationships between the tables in **kina**.

The **kina** database is implemented as a relational database; i.e., each table is a special case of the mathematical construct known as a *relation* and hence elementary relation theory is used to deal with the data within tables and the relationships between them. There are three types of relationships possible between tables, but only one exists in **kina**: one-to-many[6]. These relationships can be seen in ERDs by connecting a single line (indicating 'many') from the child table; e.g., *t_sample*, to the parent table; e.g., *t_catch_site,* with an arrowhead (indicating 'one') pointing to the parent.

Every relationship has a mandatory or optional aspect to it. If a relationship is mandatory, then it has to occur at least once, while an optional relationship might not occur at all. For example, in Figure 1, consider that relationship between the table *t_catch_site* and it's child table *t_sample.* The symbol 'o' by the child *t_sample* means that a catch site can have zero or many sample records, while the bar by the parent *t_catch_site* means that for every sample there must be a matching catch site record.

These links are enforced by referential constraints[7]. Constraints do not allow *orphans* to exist in any table; i.e., where a child record exists without a related parent record. This may happen when: a parent record is deleted; the parent record is altered so the relationship is lost; or a child record is entered without a parent record

Constraints are shown in the table listings by the following format:

```
Referential:      constraint name (attribute[, attribute])  |INSERT|
                                                            |DELETE|
                  parent table (attribute[, attribute])
```

[3] Also known as a database *schema.*
[4] A primary key is an attribute or a combination of attributes that contains an unique value to identify that record.
[5] A foreign key is an attribute or a combination of attributes that is a primary key in another table.
[6] A one-to-many relationship is where one record (the *parent*) in a table relates to one or many records (the *child*) in another table; e.g., one landing in *t_catch_site* can have many samples in *t_sample* but one sample can only come from one catch site.
[7] Also known as integrity checks.

Note that the typographical convention for the above format is that square brackets [] may contain more than one item or none at all. Items stacked between vertical lines | | are options of which one must be chosen.

For example, consider the following constraint found in the table *t_biological*:

```
Referential:        Sample no. not in t_sample (sample_no) INSERT
                    t_sample (sample_no)
```

This means that the value of the attribute *sample_no* in the current record must already exist in the parent table *t_sample* or the record will be rejected and the following message will be displayed:

```
*** User Error: insert constraint 'Sample no. not in t_sample' violation
```

For tables residing in external databases, the parent table name will be prefixed by the name of the database.

Section 5 lists all the **kina** tables as implemented by the Empress RDBMS. As can be seen in the listing of the tables, a table's primary key has an unique index on it. Primary keys are generally listed using the following format:

```
Indices:            UNIQUE index_name ON (attribute[, attribute])
```

where attribute(s) make up the primary key and the index name is the primary key name. These prevent records with duplicate keys from being inserted into the tables; e.g., a record with an existing sample number.

The database listing (Tables 1-5) show that the tables also have indices on many attributes. That is, attributes that are most likely to be used as a searching key have like values linked together to speed up searches. These indices are listed using the following format:

```
Indices:            NORMAL (2, 15) index_name ON (attribute[, attribute])
```

Note that indices may be simple, pointing to one attribute or composite pointing to more than one attribute. The numbers "…(2, 15)…" in the syntax are Empress RDBMS default values relating to the amount of space allocated for the index.

3.2 Database design

The overall database design is influenced by the duality of its function: to store commercial catch and effort data (from the SOP and where necessary, the participants); and to store biological information (from both the SOP and Marine Research).

3.2.1 Commercial catch and effort data structures

In the first case, the daily catch and effort data from catch log sheets are represented by the table *t_catch_site* (Table 1). Each record in the table contains catch and effort information by a participant, for one day's fishing at a site by diver. So if a participant has only one diver and the day's catch was from three sites, then there would be three records in this table. If a participant had two divers and each diver had three sites for the day's catch, then there would be six records in this table. The catch data (in kilograms) are held in the attribute *site_catch_wt*, and the effort data (time diving in decimal hours) are held in the attribute *site_time_diving*.

The primary key for this table is actually quite complex, being a combination of *participant*, *fishing_date*, *divers_name*, *fma*, and *site_no*. The site number (*site_no*) is only a logical identifier of a diver's dive site per day. Any one geographical site may have many logical site numbers assigned to it over time.

There are two attributes, *latitude* and *longitude*, for recording the geographical location of a dive site. However, this information is going to be limited to those divers with GPS, therefore, in addition to the latitude and longitude of a dive site, provided by the daily catch log sheet, one other attribute *site_name* has been included to increase the flexibility of catch data queries. This attribute is designed to hold the common geographical name of the dive site, e.g., the name of the bay, island, etc. It is not vital for this attribute to be used at the time of data entry, but may inserted at a much later date when necessary.

In order to enhance relationships between *t_catch_site* and other tables within **kina**, two further attributes are added. The first, *grid_ref*, records which grid reference a dive site is located. This is vital in order to link to the catch sampling data, and will be discussed in more detail later. The second attribute, *stratum*, records which Research stratum area a dive site is located. This allows commercial catches to be used along side Research data.

3.2.2 Biological data structures

Both SOP catch sampling and the Research dive sampling programmes collect biological data. Details of each sample collected produce a record in the table *t_sample* (Table 2).

All samples have a few attributes in common. Each sample collected, for both SOP and Research sampling, has its own unique sample number as defined by *sample_no*, which becomes the primary key. The date on which the sample was collected is recorded in the attribute *fishing_date*. The attribute *divers_names* records the name of the diver(s) who caught the sample. Both the SOP and Research samples are taken from within various strata as defined by the Kina Research Group, with the stratum number being recorded in the attribute *stratum*.

The remaining attributes are relevant just to the SOP samples. The attribute *observer* records the observers name; *grid_ref* records which catch sampling grid reference the sample was taken from (to link the sample back to a catch record); and *percent_samp* records what percentage of the total catch from the site was sampled.

Each sample produces a test diameter frequency that is held in the table *t_test_dia* (Table 3). Each record in this table contains a diameter class (stored by the attribute *test_dia*) and the number of kina that are in that class (stored by the attribute *no_t*).

A further 25 kina are taken from the sample for detailed biological analysis. The details of these biologicals are held in the table *t_biological* (Table 4). The attributes of this table are all reasonably self-explanatory - *whole_weight* records the whole weight (in grams) of the kina; *test_diameter* records the diameter (to the nearest millimetre) of the kina's test; *gonad_volume* records the volume (in millilitres) of the kina gonad; and *gonad_colour* records a numeric code (up to two digits) for the colour (and hence the quality) of the gonad. Note that SOP samples only have whole number codes for colour, while Research samples have decimals for increased precision.

3.2.3 Grid references

There is an additional table, *t_grid_ref* (Table 5), which has two roles in this database.

The first is to act as a coarse error check for invalid grid references. Before a catch or a sample record can be entered into their respective tables, the grid reference contained in the record must exist in *t_grid_ref* before it can be entered. Also, many of the possible grid references are invalid for such reasons as they occur completely on land or on water that is too deep for divers to reach. The attribute comment} is used to distinguish between valid and invalid grid references. A null in the comment field means a valid grid reference; otherwise the comment becomes an error message such as “Grid reference on land!”.

The second role that *t_grid_ref* has to play is to provide a link between the SOP samples and the Research samples by recording which Research stratum number each grid reference belongs to. Note that not all grid references have a stratum number assigned to them, e.g., where grid references are on land.

4 Table Summaries

The following is a listing of the tables contained in the **kina** database:

1. t_catch_site : contains catch (site location, kilograms landed, dive time, etc.) information for each grid reference fished by a diver for one participant for a day.

2. t_sample : contains details for a sample of kina made by either the SOP or Research, e.g., when the sample was collected, and who by?

3. t_test_dia : contains test diameter frequency data from a sample.

4. t_biological : contains further biological information on a sub-sample of twenty five kina, e.g., whole weight, test diameter, gonad volume and colour.

5. t_grid_ref : contains all possible grid references and comments about them. A null comment means that the grid reference is a valid one; else the comment is an error message. Allows comparisons between the SOP grid references and the Research stratum.

5 kina Tables

The following are listings of the tables in the **kina** database, including attribute names, data types (and any range restrictions), and comments.

5.1 Table 1: t_catch_site

Comment: Catch and effort information by grid reference for each diver for each participant per fishing day.

Attributes	Data Type	Null?	Comment
participant	character(30,1)	No	Name of the participating fisher.
fishing_date	date(5)	No	Date the fishing took place.
divers_name	character(50,1)	No	Name of the diver.
fma	character(5,1)	No	Kina fishery management area code.
site_no	smallint	No	A sequential number for each different site dived on for the days catch.
site_catch_wt	integer		Total catch weight (kg) of kina for a site.
site_time_diving	decimal(4,2)		Total time (hours) spent diving for the catch at the site.
latitude	decimal(5,1)		Latitude (DDMM.m format) of the dive site.
longitude	decimal(6,1)		Longitude (DDDMM.m format) of the dive site.
grid_ref	character(3,1)		3 character grid reference, to link to any sample in t_sample.
site_name	character(20,1)		General diving site; i.e., common name of the given latitude and longitude.
stratum	smallint range 1 to 6		Marine Research stratum number from which the catch was taken from.

Creator: dba
Indices: UNIQUE catch_site_key ON (participant, fishing_date, divers_name, site_no)
NORMAL (2, 15) ON (grid_ref)
NORMAL (2, 15) ON (stratum)

5.2 Table 2: t_sample

Comment: Details of each sample collected by both the SOP catch sampling and the Marine Research sampling programme.

Attributes	Data Type	Null?	Comment
sample_no	character(8,1)	No	8 character unique identifier for each sample.
participant	character(30,1)	No	Name of the participating fisher.
fishing_date	date(5)	No	Date the fishing took place.
observer	character(20,1)		Observer's name.
divers_names	character(50,1)		Name(s) of the diver(s) who caught the sample.
grid_ref	character(3,1)		3 character grid reference for the location where the kina was caught.
percent_samp	decimal(4,1) range 0.1 to 100.0		Percentage of the whole catch that was sampled.
stratum	smallint range 1 to 8		Numeric code for the geographic division of the study area.
site_no	smallint		A randomly selected study location from where kina were sampled, for Research samples only.
comments	text(240,240,240,1)		Any comments about the sample.

Creator: dba
Indices: UNIQUE sample_key ON (sample_no)
NORMAL (2, 15) ON (participant)
NORMAL (2, 15) ON (grid_ref)
NORMAL (2, 15) ON (stratum)

5.3 Table 3: t_test_dia

Comment: Test diameter frequency data for a sample.

Attributes	Data Type	Null?	Comment
sample_no	character(8,1)	No	8 character unique identifier for each sample.
test_dia	integer	No	Diameter class (mm) of the kina test.
no_t	integer	No	Total number of kina in the diameter class.

Creator: dba
Referential: Sample no. not in t_sample (sample_no) INSERT t_sample (sample_no)
Indices: UNIQUE freq_key ON (sample_no, test_dia)

5.4 Table 4: t_biological

Comment: Kina biological data for a sub-sample of the test diameter sample.

Attributes	Data Type	Null?	Comment
sample_no	character(8,1)	No	8 character unique identifier for each sample.
fish_no	smallint	No	A sequential number for each fish sampled.
whole_weight	integer		Whole weight (gm) of the kina.
test_diameter	integer		Diameter (mm) of the kina test.
gonad_volume	integer		Volume (ml) of the kina gonad.
gonad_colour	decimal(2,1) range 1.0 to 5.0 match "*.[05]"		Decimal colour code for the kina gonad.

Creator: dba
Referential: Sample no. not in t_sample (sample_no) INSERT t_sample (sample_no)
Indices: UNIQUE biol_key ON (sample_no, fish_no)

5.5 Table 5: t_grid_ref

Comment: Comments for all possible grid references. An entry in the comment field means that any dive sites referencing this grid are in error. The comment attribute is therefore the error message.

Attributes	Data Type	Null?	Comment
grid_ref	character(3,1)	No	3 character grid reference.
stratum	smallint		Marine Research stratum number which the grid reference is in.
comments	text(60,60,60,1)		Null if the grid reference is valid, else this field is an error message.

Creator: dba
Indices: UNIQUE grid_key ON (grid_ref)
NORMAL (2, 15) ON (stratum)

6 References

1. McShane, P. E. & Naylor, J. R. 1991: A survey of kina populations (Evechinus chloroticus}) in Dusky Sound and Chalky Inlet, southwestern New Zealand. N.Z. *Fisheries Assessment Research Document 91/17*.}

2. McShane, P. E., Naylor, J. R., Anderson, O., Gerring, P., & Stewart, R. 1993: Pre-fishing surveys of kina (Evechinus chloroticus}) in Dusky Sound, southwest New Zealand. *N.Z. Fisheries Assessment Research Document 93/11*.}