

GENERAL

GEOMLIB (β)

**MARINE ENGINEERING GEOPHYSICAL
DATA PROCESSING TOOLBOX**

In progress

Ivan V. Dmitriev
07.02.2020

Contents

Introduction 3

1 Geophysical data “components” 4

2 Geophysical data acquisition and processing stages 5

3 Data structures and contents 8

 3.1 Variables with fixed names 9

 3.1.1 SU (setup information) 9

 3.1.2 RS (raw sensors data) 9

 3.1.3 ML (“map-content layer” variable) 10

 3.1.4 GL (“geometry content layer” variable) 11

 3.1.5 PR (survey lines variable) 11

 3.2 Field names, postfixes and prefixes 12

 3.3 Datasets 14

 3.4 Data Content types 15

Citation 17

Tables list

Table 1 Functions sets 3

Table 2 “Navigation” fields’ names (example) 12

Figures list

Figure 1 Data acquisition and processing chart 5

Figure 2 Functions and scripts input/output variable types 8

Figure 3 Measurements’ data content types 16

Introduction

The Ge0MLib library includes algorithms for Marine Engineering Geophysical Survey data logging and processing. The follow tasks are concerned: Navigation, Sub-bottom Profiling, Side Scan Sonar, Magnetic Survey and High Resolution Seismic.

Data processing functions are developed in MatLab2015b (<https://mathworks.com/>). Data loggers are developed in Free Pascal (<http://www.freepascal.org/>). The project's site: <http://ge0mlib.com>.

The Ge0MLib presents the comfortable space (predefined fields, variable's structures, basic functions) for manipulations with geophysical and navigation data. The MatLab presents space for quick software development and compact code writing.

The specific Ge0MLib features are:

- the usable variables' structures for Geophysical Survey's tasks decision;
- vessel with equipment can be regarded as a single object described math models (for example, some corrections in navigation sensors data can easy applied for all towing/mounting equipment's position in sgy, jsf and xtf formats);
- easy access to each byte of Geophysical data;
- access to software source codes and algorithms (possibility for code modification);
- special tools for serial sensors' data logging and synchronization.

The functions sets are shown in the *Table 1*.

Table 1 Functions sets

Name	Description
gData&gFields	Row, RowM and Matrix content's functions.
gNav&gMap	Navigation and coordinates manipulations. Geometrics tasks decision, track-plots/pipelines/line-planning, picking, graphics.
gLog	Serial data logging/read/write.
gJsf	Jsf files read/write/manipulations.
gXtf	Xtf files read/write/manipulations.
gSgy&gUhr	Sgy files read/write/manipulations. HR/UHR Seismic logs and geometry.
gMagy	Magnetometers data read/write/manipulations.
gP190	P190 files read/write/manipulations.
gWfr	Images with coordinate world-files and XYZ-grid-files manipulations.
gAcad	AutoCAD scripts creation.

1 Geophysical data “components”

The measured “geophysical field” is characterized the follow components:

- time reference (synchronization) for physical field measurement;
- coordinates reference for physical field measurement (coordinates/attitude for all geophysical system’s components);
- physical field measurements’ value (includes active source/transmitter data and receiver data).

The components (in the wide sense) are characterized the **physical field distribution in time and space**. Each component is calculated/formed using a number of sensors and algorithms. There are:

- 1) The data from “**physical field sensors**”. This is geophysical equipment data.
- 2) The data from “**time sensors**”. Usually, this is navigation equipment data or computer’s clock.
- 3) The data from “**attitude/coordinates sensors**”. This is navigation equipment data.

So, the navigation equipment’s measurements are integral part of geophysical field data (its mean “physical field distribution in time and space”).

There are positioning tasks examples:

- nod’s position calculation in a vessel (used GPS, MRU/IMU, Gyrocompass, Draft-sensor, etc.);
- towing equipment’s position calculation by Layback (used nod position, cable length, depth, etc);
- off board equipment’s position calculation using USBL (used GPS, MRU/IMU, Gyrocompass, USBL data, etc.).

There are two related “time sensors” examples, were realized in ge0mlib:

- the “global” source is GPS data flow with GPS PPS (1 second discrecity);
- the “local” source is the computer clock data flow with 0.001 second discrecity (gLog tools).

On the one hand, the computer clock time is used for interpolation inside GPS’s 1-second intervals; on the other hand the GPS clock applies long-time-drift correction to computer’s clock data. The gLog software is used to link any messages from serial port to computer’s clock. The one GPS data flow with GPS PPS can be used for several computer clocks’ time reference.

2 Geophysical data acquisition and processing stages

The Geophysical Survey data acquisition and processing chart (accepted in Ge0MLib) are shown in *Figure 1*. A computer's time and GPS time with PPS are used for sensors synchronization.

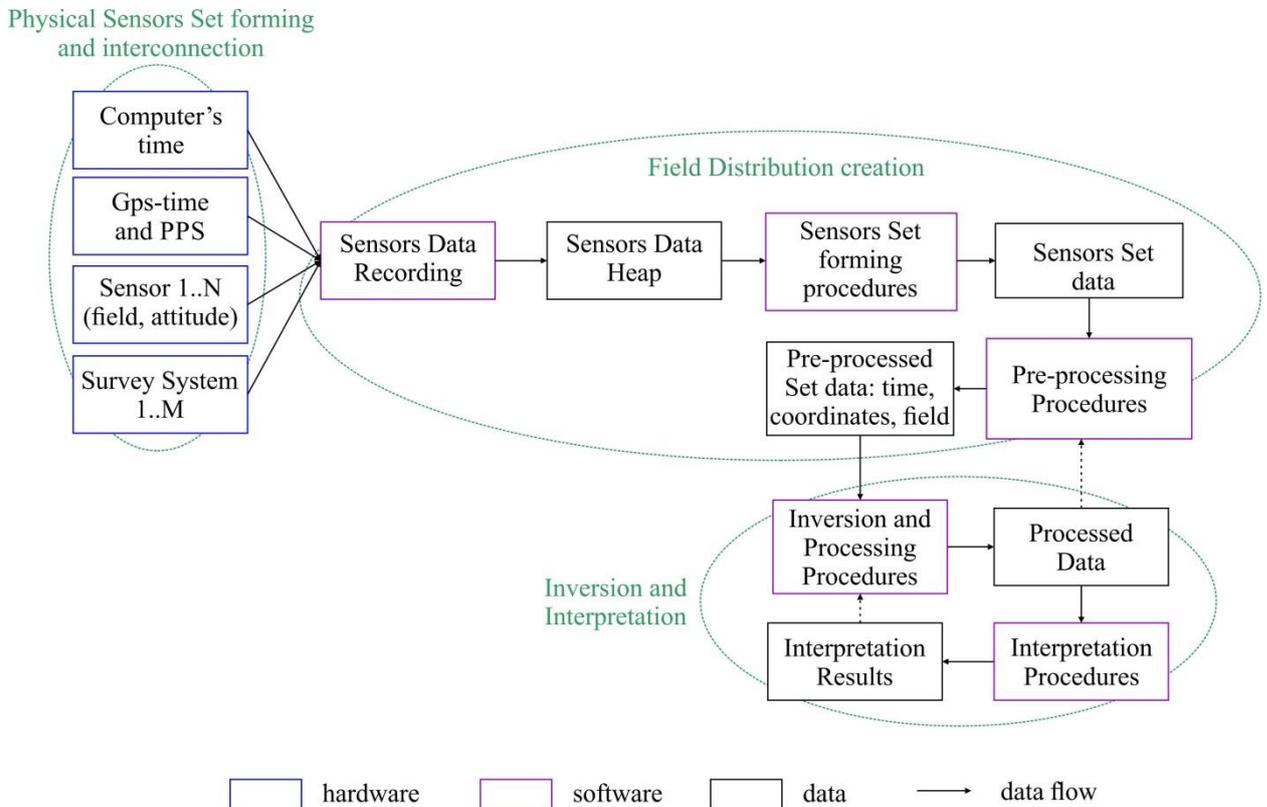


Figure 1 Data acquisition and processing chart

There are follow stages:

0) The Sensors/Hardware set forming and interconnecting (there are physical fields' sensors, timing sensors, navigation sensors, etc.) in accordance with assigned task decision (magnetic UXO survey; geological SBP survey).

1) Data acquisition/recording

The data flows from sensors come to data recording software; the data flows linkage is required. The main source for linkage is time-stamps (synchronization – sensors data link to time-of-measurements presented by “time sensors”). This stage result is Sensors Data Heap (files), which contained linked records for sensors we have in digital format; there is the “on-line survey result” or “Raw Data”.

2) The Sensors Set data forming

The Sensors Set is formed from a number of sensors. There are follow procedures for each sensor:

- a) Data import and data search (into Sensors Data Heap) using defined time (the RS variable created, – see below);
- b) The sensors data are bind to common time (computer clock or/and GPS-clock) and cut to “units”. The Sensors Set “unit” start-and-end time is the survey line record time (used Survey Log-file and PR variable created – see below).
- c) The “bad” measurements delete from Raw Data. This measurements natures is not defined by measured value physics, it is define by extraneous (third-party) forces. There are two “bad” measurements types:
 - Spikes and other “features”, which cannot defined by measured value physics;
 - Values with bad quality flag, included to data flow from sensor.
- d) Measurements smooth/filtering/de-noise procedures for normal-distributed noise decrease, sensors’ output resolution influence decrease, sensors characteristics applying (for example: noise’s frequency decrease using amplitude-frequency characteristic for seismic source and seismic receiver).

The Sensors Set forming procedures are referring to applying “specific” sensor’s characteristics; the Sensors Set “processed data” look like to “ideal sensor data” without own sensor features.

The Sensors Sets examples, for different geophysical equipment:

- SBP Set, if a transducer and receiver are “single point”;
- SBP Set, if a transducer (source) and receiver are “different points”;
- Seismic Set with single source and multi-channel receiver;
- Magnetometer/gradiometer Set (several magnetic sensors can used);
- SSS Set;
- Single-beam Set;
- MBES Set.

3) The Pre-processing Procedures and Pre-processed Set data forming

The Pre-processing procedures mathematical model deal with “sensors data combination” to create “geophysical field distribution in time and space”. The sensors “specific” characteristics was applied in previous stage; for current stage the mathematical model deal with abstract “measurements results”, which characterized measurement’s error. The model’s task example: SBP-nod’s position calculation in a vessel (used sensors are GPS, MRU/IMU, Gyrocompass, Draft-sensor, etc.).

There are follow steps for Pre-processed Set creation:

- a) Geophysical sensors coordinate (attitude) calculation on the basis of navigation sensors data;
- b) Sensors (navigation and geophysical) data interpolation to target-sensor time;
- c) Some calculations in accordance with pre-processing geophysical processes models (for example: Earth’s magnetic variations or deviations applying);
- d) Disambiguation for Sensors Set data (measurements or model correction);

- e) Using Sensors Set, create the Pre-processed Set (PR variable fields – see below), which include: physical field measurements, time reference for physical field measurements, space/coordinates reference for physical field measurements.

4) The Processing Procedures and Processed Data forming

This procedures deal with physical field distribution in time and space (Pre-processed Set). The procedures' mathematical model deals with physical field equations. The procedures' tasks are inverse problem decision or the same. The Processed Data is the physical field's sources distribution/characteristics or another inverse geophysical tasks' decision.

5) The Interpretation Procedures and Interpretation Results forming

The Interpretation Procedures is the Processed Data “interpretation” in the terms of physical characteristics and objects for geological/engineering knowledge area. The Interpretation Results are the some information about geological/engineering “features” for survey area.

The steps 3-to-4 and 4-to-5 can be recursive for results best conformity to selected mathematical models.

The Ge0MLib area is the stages 1, 2, 3 and to a lesser extent stage 4.

There are follow **Data Abstraction Levels**:

- 1) The Sensors Set forming procedures deal **with “specific” sensor’s characteristics**. All “specific” information is applied to data this stage.
- 2) The Pre-processing Procedures deal with “ideal sensors” without specific features and without irrelevant noise. The “ideal sensors data” used for Pre-processed Set creation – **physical field distribution** calculation (there are physical field values, time and space coordinates).
- 3) The Processing Procedures deal with physical field equations. The result is inverse **geophysical tasks’ decision**, named Processed Data.
- 4) The Interpretation Procedures deal with transfer “information” from physical field’s knowledge space to **geological/engineering knowledge objects** space (create Interpretation Results).

3 Data structures and contents

The data, geOmlib's functions used, can be considered as a specific content by the data-structure and "sense load". The content's types are described below; there are: Information content; Row-content; RowM-content; Matrix-content; MatrixM-content.

The function and scripts input/output can characterize with the several levels of formalization (*Figure 2*):

- The first level contained the "universal" functions with input/output parameters defined as a vectors, numbers, strings, etc.;
- The second level is a "structures with predefined field names". The field-names are define "sense load" of data contained (coordinates, heading, temperature, etc.). The functions have "knowledge" about field names, make some manipulations with data in fields and create new "known" fields
- A number of same "structures with predefined field names" can formed a vector-structure. For example, a number of sgy-files (SBP survey) or P1/90-files.
- The third level contained functions with additional input variables – a field names will use (if indeterminacy for predefined field-names take place) or prefixes/postfixes will added to predefined field-names. There are special rules for prefixes/postfixes using.
- The fourth level of formalization is characterized the several cell-vectors with fixed names. There are: SU (project's setup information), RS (raw sensors data), ML ("map-content layer" variable), GL ("geometry content layer" variable), PR (survey lines) and others. Each cell contained its own object (sensor's measurements, survey line data, map-object for drawing) with own second or third level structures.

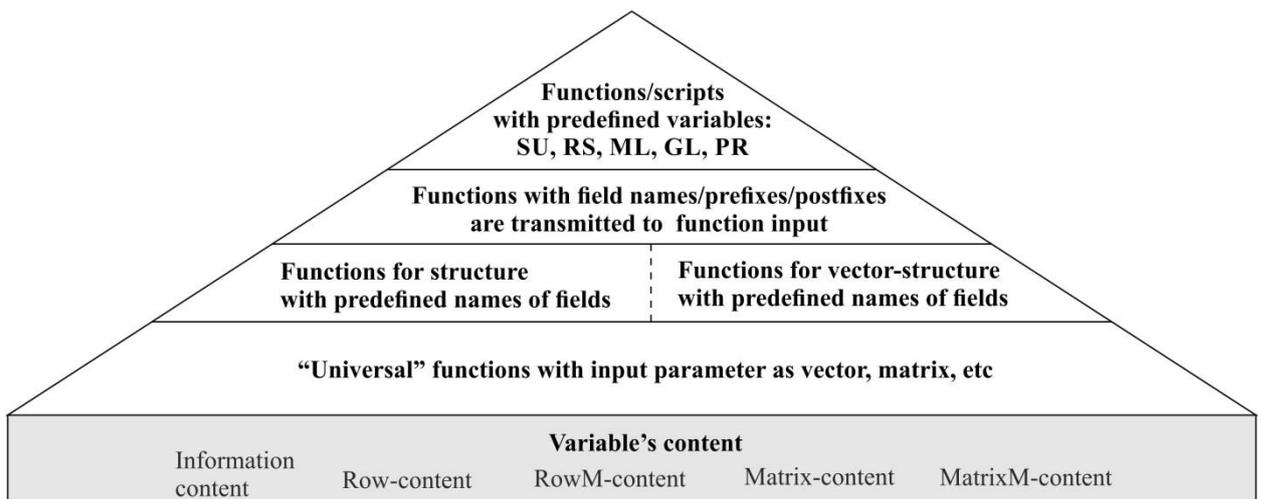


Figure 2 Functions and scripts input/output variable types

3.1 Variables with fixed names

There are follow cell-vectors variables with fixed names:

SU – project’s setup information;

RS – raw sensors data (data heap);

ML – “map-content layer” variable;

GL – “geometry content layer” variable;

PR – survey lines variable (includes sensor’s data interpolated to target-sensor time for each survey line).

Each cell contained its own object; there are follow usually used “objects” examples:

-- Survey Log-file – RS{SU.LLog};

-- Line planning (map-layer) – ML{SU.LPlan};

-- Infrastructure objects’ coordinates, coastline, wells head coordinates, etc. (map-layer) – ML{...};

-- Pipelines or cables coordinates and other characteristics (map-layer and geometry-layer) – GL{...};

-- Bathymetry data / digital relief model’s data for different years (map-layer and geometry-layer) – GL{SU.Bott }, GL{...}.

3.1.1 SU (setup information)

The “setup information” variable includes information which concerned global project’s settings.

The variable fields example:

SU.RootDir='d:\2019\009_Grad'; – the name of project’s root folder;

SU.LLog=1; – RawSensor's cell number with Survey Log-file;

SU.LPlan=1; – map layer’s cell number with Line Planning;

SU.Bott=1; – geometry layer's cell number with actual sea bottom;

SU.NavS=struct('EllipParam',[6378137 sqrt(2.* 1/298.257-(1/298.257).^2)],'EllipTransParam',[6.102269 -10.161 -5.186 -0.050 0.0701 0.0583 2.0293e-6], 'EllipForvTransFunc','gNavGeoc2Geoc1032', 'EllipRevTransFunc', 'gNavGeoc2Geoc1032inv', 'TargCode',2); – Navigation Sensors coordinates datum;

SU.NavP=struct('EllipParam',[6378137 sqrt(2.* 1/298.257-(1/298.257).^2)], 'ProjParam',[0 121 0.9999 250000 0], 'ProjForvFunc', 'gNavGeog2ProjUtm', 'ProjRevFunc', 'gNavProjUtm2Geog', 'TargCode',6); – Project’s coordinates datum;

SU.CompTimeLocShift=8*3600; – local time shift in seconds;

SU.MagNorm=[44000 -3.60 35.50]; – magnetic locally-normal field [T(nT) D(deg) I(deg)].

3.1.2 RS (raw sensors data)

The “raw sensors data” variable is the cell vector; it is includes “information” which necessary for sensor’s files loading and the results of loading. The “information” is the different for different sensor’s types. The variable fields example:

1) The information for “Magnetic survey Log-file” loading (sensor’s number SU.LLog=1)

RS{SU.LLog}.fRec='Log_mag.txt'; – file with Log (located in the project’s root folder);

RS{SU.LLog}.fRecFormat='gLLogShot'; – format identification code;

RS{SU.LLog}.fgLogDelim=[]; – un-applicable;

2) The information for “Towed body position” loading (sensor’s number 2)

RS{2}.fRec='000_TB\'; – the name of folder contained files with towed body position;

RS{2}.fRecFormat='gLogGPGGA'; – position’s format identification code;

RS{2}.fgLogDelim=['~' ',']; – gLog delimiters for Computer Time label (the data flow was recorded using gLog);

RS{2}.Period=1; – sample period (NaN if not applicable).

3) The information for “SeaSpy2 magnetic gradiometer data” loading (sensor’s number 2)

RS{7}.fRec='000_Mag3\'; – folder contained files with magnetic gradiometer data;

RS{7}.fRecFormat='gLogSeaSpyGrad'; – magnetic gradiometer identification code (examples: MagLogG88Magy, MagLogG88Grad, gLogSeaSpyMagy, gLogSeaSpyGrad);

RS{7}.fgLogDelim=['~' ',']; – gLog delimiters for Computer Time label (the data flow was recorded using gLog);

RS{7}.Period=0.5; – sample period;

RS{7}.ID=[999 999 999;999 nan 999]; – serial numbers [magnetic_sensor1_SN, altimeter1_SN, depth_sensor1_SN;; magnetic_sensorN_SN, altimeterN_SN, depth_sensorN_SN]; the serial numbers are used with two aims: (1) the equipment identification; (2) the detection if the equipment installed or not (nan value);

RS{7}.Coef=[1 0 1 0;1 0 1 0]; – scale and bias coefficients for Altimeter and Depth sensor [Altimeter1Scale Altimeter1Bias Depth_sensor1Scale Depth_sensor1Bias;; AltimeterNScale AltimeterNBias Depth_sensorNScale Depth_sensorNBias].

3.1.3 ML (“map-content layer” variable)

Map-content layer includes data will used for visualization only (look like the MaseMap); it is vector-data (usually in PolyLine structure – survey lines planning, infrastructure objects position) or some geo-referenced images (bathymetry image, SSS mosaic, Backscatter image). The “Map-content layers” data can be drawn in MatLab or exported to software same to AutoCAD. The variable is the cell vector; it includes “information” which necessary for “map-layer” files loading and the results of loading.

The variable fields example:

The information for Line Planning loading

ML{SU.LPlan}.fRec='LinePlan20190916.txt'; – the name of file with Line Planning (located in the project’s root folder);

ML{SU.LPlan}.fRecFormat='LinePlan'; – format identification code;

ML{SU.LPlan}.NDraw=2; – draw order (NaN if not drawing).

3.1.4 GL (“geometry content layer” variable)

Geometry-content layer includes data will used for some calculations; it is vector-data (for example – pipeline position, depth and radius will used for calculation crossing with survey line or seismic hyperbola modeling for real track) or some grids (for example – bathymetry data will used for seismic bottom leveling). The variable is the cell vector; it is includes “information” which necessary for “geometry-layer” files loading and the results of loading.

The variable fields example:

The information for Bathymetry data loading

GL{SU.Bott}.fRec='3961_EGS#2.pts'; – the name of file with Bathymetry data (located in the project’s root folder);

GL{SU.Bott}.fRecFormat='Xyz2Triang'; – format identification code.

For drawing the geometry-layer must be converted to “Map-content layer”.

3.1.5 PR (survey lines variable)

The “survey lines variable” variable is the cell vector; it is includes sensor’s data from RS (raw sensors data) which interpolated to target-sensor time. Each survey line is cut using start-survey-line time and end-survey-line time from Survey Log-file.

3.2 Field names, postfixes and prefixes

A sensors data, imported to MatLab includes the predefined field names for each type of data. The predefined coordinates and other “navigations” fields for some point are shown in *Table 2*. These fields’ names will take place for **RS** variable when sensor’s data will read using “data-reading functions”; these fields’ names are used and modification (with postfixes and prefixes) in **PR** variable. The universal field names let to use field-structures with different functions, which are to know “universal field names”.

Table 2 “Navigation” fields’ names (example)

Field name	Field Description
GpsLat	Latitude/Easting (WGS84 usually); DD.DDDDDDD.
GpsLon	Longitude/Nording (WGS84 usually); DDD.DDDDDDD.
GpsAltSea	Height by Gps, GpsAltSea data
GpsHgtGeoid	Height by Gps, GpsHgtGeoid data
GPSHeading	True heading by differential_GPS/Gyrocompass/MRU.
GpsE	Rectangular projection Easting.
GpsN	Rectangular projection Northing.
GpsH	Height created when Ellipsoid-to-Ellipsoid coordinates transformation take place; field used for coordinates transformation’s stability
GpsKP	Point’s ID, it can be measurement number, kilometer point’s number or any other.
GpsZ	Points Z-coordinate for vertical geodetic datum (pipeline or towing equipment position).
Altitude	Point’s Altitude (usually by altimeter for towing equipment)
Depth	Point’s Depth (usually by depth sensor for towing equipment)
Heading	Heading in Rectangular projection.
WaterDepth	Water Depth for current points coordinates (can be “actual” or referenced to some system MLS, LAT, Baltic, etc).
<u>XXXXRaw</u>	Raw data (no smoothing, de-spiking, etc), where XXXX is any field name.

The **RS (raw sensors data)** variable field-names include the special **postfix “Raw”**; it is used for raw-original data keeping (which was loaded form sensor’s files). For example, the original coordinates from GpsE and GpsN can be kept in GpsERaw and GpsNRaw, after that, the coordinates in GpsE and GpsN fields can be de-spiked and smoothed.

The **PR (survey lines)** variable’s cells include different sensors data (interpolated to target-sensor time) and “calculated” data. The **prefixes** are used to specify mentioned data’s field names. There is follow rule:

The different sensors which includes fields with similar names are assign the own “field-names user-defined prefixes”. For example, tow-point coordinates (with fields, named as GpsE, GpsN and GpsH for data was loaded) interpolated to magnetic field measurement time can be assigned the prefix TP (TPGpsE, TPGpsN and TPGpsH); for towed body coordinates can be assigned prefix TB (TBGpsE, TBGpsN and TBGpsH); for calculated magnetic sensors (located on the towed body) coordinates can be assigned prefix TS (TSGpsE, TSGpsN and TSGpsH). So, one survey line cell can include tow-point, towed body and sensors coordinates.

The target-sensor original (not processed) data (for **PR** variable) can kept in the fields with names includes **postfix “Raw”**. The sensors data (interpolated to target-sensor time) and “calculated” data usually cannot have postfix Raw; there can be some exceptions if we apply some processing to interpolated/calculated data and we need to keep pre-processed “original” data.

3.3 Datasets

Some sensor data, loaded from a number of files to variable-structure with predefined field names, can be offer as a Dataset. It is mean that each survey line loaded to own vector-structure element.

The example for vector-structure A:

```
-- A(1..n).GpsLat;  
-- A(1..n).GpsLon;  
-- A(1..n).GpsAltSea.
```

The Dataset used for several files processing; for example, we can load a number of P1/90 files to Dataset and create track-plot map for survey lines.

There can be more difficult Dataset structures. For example, we can load a number of Sgy-files to Dataset. Sgy-file read in MatLab to three variables, usually named SgyHead (binary, textural and extended textural headers), Head (trace headers) and Data (Matrix with trace's data). The Dataset is formed as

```
-- SgyHead(1..n)  
-- Head(1..n)  
-- Data{1..n}.
```

SgyHead(m) and Head(m) are the vector-structure element with predefined fields. The Data{m} is the cell-vector element which contained seismic traces data or tmp-file name with seismic traces data (see gSgy manual).

So, we can make uninterrupted SBP survey (not stop survey between survey lines, not change survey parameters); the Sgy-files will cut on-line software by size. We can load Sgy-files to Dataset to **RS (raw sensors data)** variable and create "profiles" (**PR** variable) formed the new Sgy-files with a vessel straight moving and run-in.

3.4 Data Content types

There are general features for measurements' data content types or "containers" (*Figure 3*):

Information-content

This content contained "constant" information about sensor's characteristics, measurements mode, etc. This information is applied to data block unit: Data Heaps' sensor, all Data Set or one survey lines' data. The information-content examples: Ellipsoid characteristics, coefficients for depth sensor or altimeter.

Row-content

This content is contained measurements results along time axis (binding/linking source axis). All rows are designed as variables' fields by usability. The row-content fields for data block unit (for example, one survey lines' data) have equal length. Row-content field examples: Easting coordinate, Northing coordinate, Water temperature, Pitch angle, Total magnetic field. There are several predefined Row-content field names for general data types, which used by GeOMLib functions.

The row-content is the equipment measurements result with "scalar output", when one measurement (along binding/linking source axis) is one number by one sensor. There are any time-series "scalar" measurements: GPS data, IMU/MRU data, magnetic sensor data, etc.

RowM-content

If we have a number of the same and synchronized sensors with row-content, it is reasonable to create matrix with a number of row. This matrix is alternate to several Row-content fields. The RowM-content example is magnetic measurements with gradiometer; there are a number (from 2 to 7) synchronized and the same magnetic sensors.

Matrix-content

This content is contained measurements results along time axis (binding/linking source axis); each measurement is time-series inserted to matrix column. The matrix-content is the equipment measurements result with "time-series-output", when one measurement (along binding/linking source axis) is the series numbers along time axis (by one receiver).

The matrix column examples are SBP or SSS traces. Other measurements (column/traces) characteristics (coordinates, time stamp, etc.) are contained in associated Row-content structure. The matrix column number and Row-content field's length is equal.

MatrixM-content

If we have a number of receivers with matrix-content, it is reasonable to create additional axis with a sensor number; we will have 3-axis dimension. A "number of receivers" example is digital-seismic streamer. There are dimension axes: 1) short time or short numbers; 2) record-time for hydrophones or recorded digits number; 3) receiver numbers.

Other measurements characteristics (coordinates, time stamp, etc.) are contained in associated Row-content structure (for example, seismic source coordinates) or RowM-content structure (for example, streamers' hydrophones coordinates).

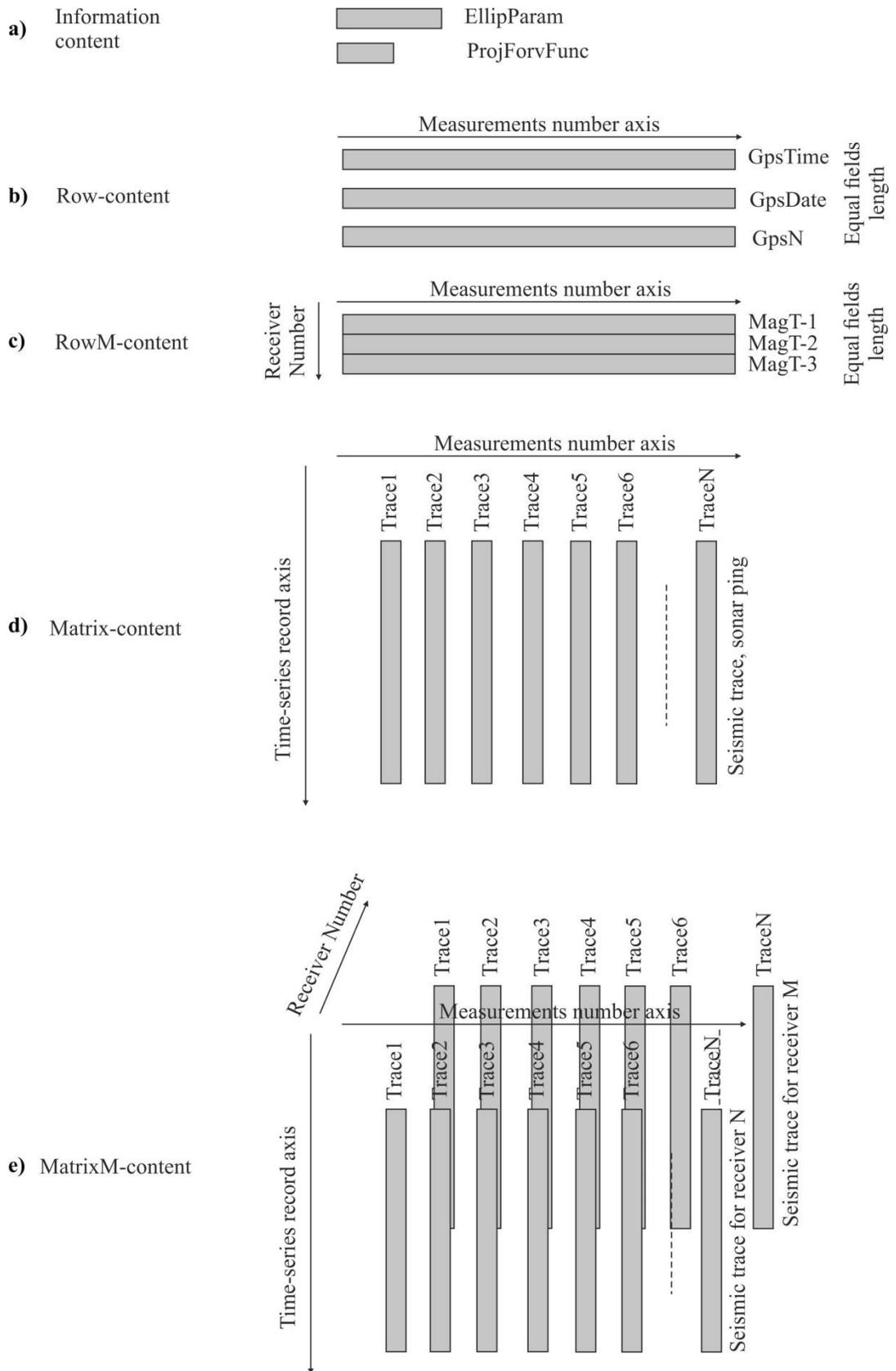


Figure 3 Measurements' data content types

Citation

1)