

Exchange format and QC estandar routines

SUMMARY

Sea level data has been measured and banking from a long time. The people responsible for the gauge operation have taking much time to calibrate the gauge in order to obtain the continuity and unicity? of the time serie. Because from a long time the data has been sending to the PSMSL the sea level references took a relevan role.

Since last two decades with the developement of the measurement systems and computer and the relevant of the study the global behavior of the sea level. The data are intercahnge between institution and used from diferrent groups. For this reason the data need to be qualify with standars procedure and metadata has to be included: The source institute, the units, the location, the equipement systems, the system reference and the reference level, etc.,. Also, in order to do not delete erroneus or suspicious values the data should be flag and any relevant comment would be included.

Because most of the tide gauge network have operated under the responasability of hidrographic and oceanographic institute and because the sea level is an oceanographic variable, the IOC have been taking an importan role in the sea level studies nad the sea level observation systems requirements, like GLOSS, standard format for interchanged and QC procedures.

One of the commonly package software used is the one developed by the jalc under auspice of the GLOSS, based on the hourly residuals and diferenced of the cosed stations visualitations.

But recently the data are register in near-real time at higher frecuency, 1,5,10 minutes intervale, and for fast delivered them estándar procedures to qualify the dat at these step has to be updated. By the way some global program like WOCE have experience in the data archaiving and QC. In the EOSS Position Paper this information has been taken into account.

The consortium of the ESEAS-RI have to built a good quality data set. Them we need to Agree and proceses the same quality procedures and for interchang to have a coomon format.

This report will folow the recommendation on the GETADE exchange format and it will take as much as posible the international existing agreement on exchange format and quality control standards procedures as

JASL/WOCE, BODC, MEDATLAS, WOCE quality control guidelines for water level.

1.- DATA & INFORMATION

Appart from the sea level data an essential additional information is nedded not only for archaiving and quality control but esential when the data are exchange and/or integrated in a regional or global data set. As it is mentioned in [IOC Manuals and Guides No. 14: Volumes I, II and III \(available at www.pol.ac.uk/psmsl/manuals/\)](http://www.pol.ac.uk/psmsl/manuals/), the information requiered to accompany the data in order to have a sea level data bank fully effective:

“

Data sets

At the tide gauge station it is advisable to register in addition to sea level heights, some related variables as atmospheric pressure, water temperature, wind direction and speed, etc. and it is essential to get information on data calibration and reference level variations. The observed data are processed routinely in order to get mean sea levels. The type of data commonly archived in the data bank are :

- *Data Observed*

- Sea level and related variables at sampling intervals
- Calibration data
- Reference levels (from levelling or GPS)

- *Data processed*

- Hourly heights, predictions? and residuals?
- Extreme values
- High and low waters, tidal ranges

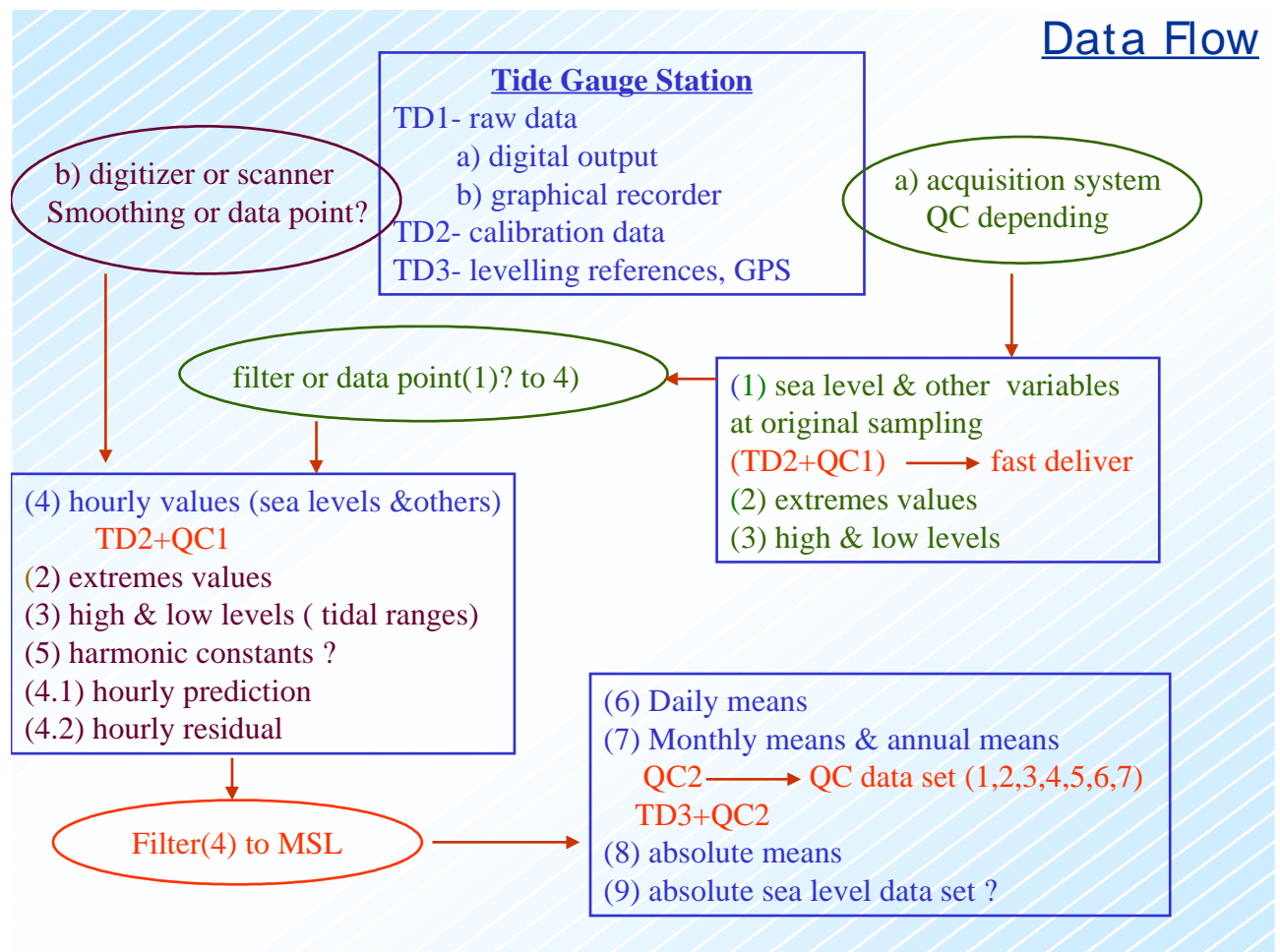
- *Data Calculated (Means)*

- Daily means
- Monthly means
- Annual means

- *Final Data Set*

- Absolute values (means and/or heights)

Data flow Scheme



2. DATA ORGANIZATION & EXCHANGE FORMAT

As mention above there are a lot of information that should be accompanying the sea level data. For this reason, special attention should be paid in the data organization. Because of the different system and facilities that the operational and data centre could have, different option are presented:

- Option 1: ASCII files for data and accompany information
- Option 2: Text files for information & ASCII files for data
- Option 3: Text files for information & RDB tables for data
- Option 4: RDB for data and information

But some advise are given by the IOC Tecnical Expert [Formatting Guidelines for Oceanographic Data Exchange](http://www.ices.dk/ocean/formats/getade_guide.htm) http://www.ices.dk/ocean/formats/getade_guide.htm .

“ Formatting Advice

Country/Platform

The country of registration of the platform (e.g. ship) and identity of the platform. It is advisable to use the IOC country and platform code tables to indicate the country and platform.

Data

Data should be reported in standard SI oceanographic units. Data may be in the same file as header information or may be stored in separate files.

Date

This conveys information about the year, month, and day on which the data were collected. A date in the form YYYYMMDD is preferred

Time

Time should be reported as hours, minutes and seconds or decimal equivalents, in UTC (GMT). Always use a 24 hour clock with hour values between 0 and 23. Whenever possible, it is advisable to convert local times to UTC.

Instrumentation

Information about what instrument was used to make the observation should be included with the data. If this is not possible some other means of identifying the method by which the observation was made should be included.

Missing Values

The use of missing value indicators is strongly recommended. There are many ways to indicate a value is missing but care must be taken to ensure that they cannot be confused with valid data. Indicators should be clearly described in the file header. **In our case there are two possibilities:**

- a) missing data: default values
- b) missing data: no data record (if there is no any type of data)

Sounding

The inclusion of sounding is a useful check of the position of a station. In the case of the permanent station is no so important but for the boyes or moored station are necessary

Station Number. (PSMSL Reference)

One of the most importan fields when data are organized and trasfered is the reference, in this case, for example the PSMSL reference should be used for the permanent sea level tide gauge station. For the non permanent stations, some criteria should be defined ”

Format Specification

Acording to the GETDATE specification for data exchange, there exist some format specifications. One very widely employed is the JASL/WOCE <http://kpau.soest.hawaii.edu/uhs/c/jaslpr2/slman2.html>, but which refers only to hourly values and daily and monthly means, The other formats that we should considered are:

BODC/WOCE:

http://bodc.ac.uk/cgi-bin/frame?http://www.bodc.ac.uk/projects/wocedac/del_mod/format.htm

MEDATLAS:

<http://www.ifremer.fr:582/sismer/program/mater/formats.htm>

The MEDATLAS format has not been describe specifically for sea level but for the profile station and time series. In this format, the data and information structure are organized in 3 parts:

Agency and platform Information

country and organisation
originator identifiers
start and end dates

Station information

Station reference
geographical position
measurement interval
Instrumentation (manufacture, calibration methods, history,etc)
Site (location, benchmark, datum relationship & history)
Processing procedures

Data parameters

parameter codes
date
parameter/s data values

But for any format/s decided the standard codes and the sea level glossary should be used.

station reference (PSMSL) <http://www.pol.ac.uk/psmsl/>
data dictionary (GF3) <http://www.ices.dk>
sea level glossary <http://www.pol.ac.uk/psmsl/>

And for the data table/file a common structure should have :

- 1) header line with codes
 - 2) data
- heights at any sampling and daily mean
 - date: yyyyymmdd
 - hour: hhmmss
 - variable(1): (Units International System)
 - flag(1): according to the codes
 - variable(2): (International System Units)
 - flag(2): according to the codes
 - cycle : (*)
 - Monthly and annual mean
 - date: yyyyymmdd
 - hour: hhmmss
 - mean : (Units International System)
 - flag: according to the codes
 - number of days used in the computations

- tide ranges data:
date, time, high value, date, time, low value, tide range, QC flag,
- extremes data:
date, time, extreme value, code, date, time, extreme value, code, QC flag
where code (1= maximum, 0= minimum)

(*) for the missing data the discussion on the default or not record the number of the cycle should be included.

But for any format/s decided the standard codes and the sea level glossary should be used.

station reference (PSMSL) <http://www.pol.ac.uk/psmsl/>
data dictionary (GF3) <http://www.ices.dk>

3 QUALITY ASSURANCE

The error that can arise in the sea level data and related parameter could be random errors, for example, transposition of numbers in manual recording or recording of observations in the wrong column, or can arise from electronic noise in measurements, problem in the communications, sensor calibration, etc. But also could be systematic errors, the inhomogeneity type errors arise primarily when there is a change in observational practices, for example, change the instrumentation or location that can result in sharp discontinuities, or even the environment surrounding of the station, as harbour constructions, land movements, etc., that can lead to a trend.

The process of quality control consist in perform various controls-checks and depend on the result change the value by interpolation or any other procedure or even leave the data unchanged but in both case flag the data according to a standard QC codes.

QC Codes

There are many schemes used to flag data quality. All are acceptable provided the interpretation of the flag is documented. One scheme employed for data reported in real-time is illustrated here. It uses a one character field with the following interpretation:

- 0 = data are not checked
- 1 = data are checked and appear correct
- 2 = data are checked and appear inconsistent but correct
- 3 = data are checked and appear doubtful
- 4 = data are checked and appear to be wrong
- 5 = data are checked and the value has been altered

In the case that data values are altered as a result of quality control practices, it is strongly advised that the original value be preserved with the data.

3.1 Preliminary Analysis

This report are not focus on the data analysis but some “highlines?” are given because to perform the data quality control some data analysis are required.

Filters:

The data normaly are registered at diferent intervale from 1 minites up to 1 heure, the most common is 5-10 minutes, only in regions where sheiches occur frecuently the data are register at 1 minites,. Appart from the convenience on keeping some of this signals produced for whether effects. It will be necessary to obtain hourly values and daily means. By filtering process some signal (low frecuency, high frecuency,etc) are eliminated depending on the frecuency cut-off. PUCH, (1987) Tides, Surges and Mean sea-Level describe useful filters that can aply to the sea level data at intervale of 5,10,15 minutes to get the hourly heights to keep the tidal phenomena, an also low-pass filter to eliminate diurnal and highest frecuencies in the hourly values and get the daily means. All these filters are simetrical and the filtered value is calculated by the expresion:

$$X_F(t) = F_0 \cdot X(t) + \sum_{m=1}^m F_m [X(t+m) + X(t-m)]$$

Where the factors F are listed in the book. (*) **ROUTINE (P6)**

In Godin (1972) there is an extensive discussion on the tidal filter. The expresion below can be apply to any time series at any interval because the weits are calculated depending on the intervale and the cut-off frecuency.

$$x'_t = \frac{1}{n} \sum_{i = -\frac{n-1}{2}}^{i = \frac{n-1}{2}} x_{t+i}$$

A test have been performed by PE and IEO with both routines, and the data result from applying the second procedure give a smoothed tidal amplitude. In fact, FOREMAN in its tidal analysis software has a procedure to recover the amplitude.

Harmonic Analysis

This analysis has to be perform in order to obtain the harmonic constituent not only for tidal prediction but also it is a very powefull tools for quality control when we inspect the residuals. Just by visualization of the residual many error are detected mainly outlier and datum shift and time shift. Also from the residuals it is posible to interpolate the value detected as error and also filling some gaps.

There exist more than one methods to obtain the harmonic constituens of the tides.;

- a) least square method analysis. FOREMAN Software.
- c) filter method analysis. ETDEVA software?

From the constituents, the heights and high low levels are calculated. Then the residual are the differences between observe and predicte data.

Example to visualize tipical errors.

Statistics

In order to validate the data, information on general behaviour of the station or grouped into oceanographic and/or climatic region needs to be calculate from historical data. In particular some statistic should be performed to obtain broad ranges means and standard desviation to perform some quality checks.

- upper and lower limits for range check.
- tidal ranges at each stations for time-continuity check.
- means and standard deviation for MSL and other variables.
- monthly extremes and standard deviation.
- others

(*)Broad ranges and standard desviation tables for the sea level and related parameters have to be prepared.

Interpolations:

During quality control, the technique to replace the spikes for an apropiate value is the linear interpolation as it is done in the JASL software not only for spikes but also for filing short gaps.

Method: Linear interpolation from prediction data

x_1, x_2 : observe data at the ends of the gaps

y_1, y_2 : prediction data at the ends of the gaps

$r_1, r_2 = x_1 - y_1, x_2 - y_2$ residual data at the end of the gaps

$$r_i = r_1 + (r_2 - r_1) \cdot ((t_i - t_1) / (t_2 - t_1))$$

$$y_i = y_1 + r_i$$

Considerations: gaps < 24 hours

$$r_i = r_1 + (r_2 - r_1) \cdot ((t_i - t_1) / (t_2 - t_1))$$

Appart from the linear interpolation ,probably other type of interpolation can be discribeif nessary.

Correlation

Some of the chek are performed by comparing the value with data from other station or from data prediction or related variables. In this case the correlation analysis are very usefull because gives not only the correlation coefficien but also the lag of maximum correlation between two series. The correlation ROUTINE can be apply to:

observed/prediction hourly to detect and correct spikes and timing error?
 Msl(1), heights(1) /other variables (pressure, temperature, msl(2),height(2) for QA
 modes / variables for QA ?

This analysis are also useful to **filling gaps (msl)**:

Procedure: case (1) xi (One reference series), Xj (original series)

Normalised series

Substitute Xj for xi

case (2) xi (Various reference series, Xj (Original series)

$$X_j = \frac{\sum_i \rho_{ij}^2 \cdot X_i}{\sum_i \rho_{ij}^2}$$

where Xi is the corresponding normalised data at the other two stations, and ρ_{ij} is the correlation coefficient between the series.

3.2 Quality Check

As mentioned above the error can arise from different causes then it will be necessary to find out all the problem. A complete QC will be needed to perform it in at least 3 etages: check metadata, check the data, QC Assesement.

- **QC0: metadata**

Format

latitude: degree (-90-90), min(0-60),sec(0-60), hemisphere: N (0-90), S(0-(-90))

longitude: degree (-180-180), min(0-60),sec(0-60), hemisphere: W(0-(-180)), E(0-180)

bottom sounding: compare with references. Etopo5 ?

if bottom sounding out of the regional scale flag =4

if the sounding is whithing (-10%) and maximun (+10%) of the 9 reference value, flag= 1 .

if outside of this intervale, flag =3.

location and bottom sounding (map visualisation)

code parameter and units: compare from GF3 codes table

dates: month (1-12), day(1-corresponding month), dates: between start date and end date

others

duplicates (period cross reference, cross position and agency)

QC1 checks the data

— Dates: in consecutive order. Including gaps?

In a time serie the mandatory parameter is the date-time, any other parameter has sense without the corresponding date-time. Sometime the series have gaps that could be from some data points to days and months or years. For this reason the check should be taking into account as much as possible those situations.

Method: check consecutive date with the calendar routine and the cycle?

Action : a) if dates are lagged just them move dates automatically. Not error, flag = 1

b) if other problem, change manually. Not error, flag = 1

Stability: Reference changes Measurement system stability & calibration practices

Improper calibration leads to reference level shifts. Most shifts are readily identifiable in hourly residual plots. They can also be seen in plots of the daily and monthly data and in plots of differences of daily or monthly values with nearby tide stations or with redundant sensors at the site in question. If a shift is identified, the proper means of correction is through the analysis of the tide staff readings and the corresponding tide gauge values. Inspect pair values. If the tide staff readings were accurate and the tide gauge was functioning well, then the points on the graph will be nearly linear with slope equal to 1.

As a rule, the UHSLC considers changes greater than 1.5 cm as significant and worthy of an investigation into the level stability.

Method: if there is information on the reference change value, add this value to the for the corresponding period.

Action: $x_i = x_{i+}$ value (apply to the time period) & add information to the comment line

Broad Range.

Comparing a measurement with established upper and lower limits that are station dependent, although some may be grouped into climatological regions. Some “seasonal” or longer terms limits may be applied. The defaults are set 3σ from mean values and the adjusted for record high and low values for stations within the area. Seasonal limits are usually set at 2σ from the mean climatological value for a specific area and month. A broad range table should be establish by station or region and by season if necessary

Method: if data is out of the regional scale (max & min from statistics)

Action a) flag 4

b) interpolate data from prediction or neighbourhood stations & and flag 5

Spikes. Time-consecutive check. Delta check. The aim is to check the change in time of a particular variable. The outliers is any value higher than the tolerance

Method: check the consecutive data with the tolerance

if (abs ($x_i - x_{i-1}$) > Tolerance

where $\delta = 2 \cdot \Pi \cdot \text{amp} \cdot \text{inter} / 720$ (tidal data), (amp from statistics or from prediction)

$\delta = 0.58 \cdot \sigma \cdot t^{1/2}$ (for non tidal data) ?

Action: a) flag = 4

b) Interpolate data as above and flag = 5

Shift and drift time.

This type of error appears mainly because of the clock malfunctions. In this case a drift time error arise but sometime the operator does not setup the time correctly and then a shift time error appears.

Methods (1) Can be detected by visualization of the residuals.

(2) Correlation (with lag) between observed data and predictive or neighbourhood station data. In case of the shift time error, the correlation have to be perform for the periods before and after the shift. In the case of drift time the correlation should be perform day by day. (other procedure)

Action: Case a) if shift is a multiple of the time interval move data according to the lag time

Case b) shift is not multiple of the interval, then it will be necessary to make the interpolation (describe the method.)

Case c) In case of the drift time after detected the times it will be necessary to interpolate (describe the method)

- “ Predicted value check: check the difference between the primary water level value relative to station and the predicted value relative to station datum against specified tolerance limit sets in SIF.
- Neighbouring check: Check the difference between the primary water level value for a station relative to the appropriate datum and the corresponding values for 2 neighbouring stations to see if a specified tolerance limit is exceeded. Values for the neighbour check are located in the separate neighbour station table.”

QC 2 Land Stability. Absolute sea level values

The land stability is one of that causes that can lead to a shift or drift datum but because there is information on that at each station (levelling or GPS monitoring benchmark) and also by models, an additional process & or check will be performed.

PGR: Remove the Post Glacial Rebound

Action: de-trend the postglacial rebound ($y_{corr}(t) = y_{orig}(t) + s \cdot t$)

where s is the trend from Peltier
document in the comment line

Absolute Data: Remove the land movement from GPS or levelling

Method: ?

Others ?

- QC Assesment:

Shift and drift data

When working with the historical data, even if the station are well documented, probaly we do not have control sheets to perform the above quality check “reference changes”. Furdermore, changes in the instrumentation, in the environment surrounding the station can result in sharp discontinuity (shift datum) or can lead to trend (drift datum). In this case some additional check should be perform to obtaine a unique reference. The shift and drift data error can be detected by visualization of the mean sea level of one station and better as diferences of two stations.

Methods: (Differences between stations, Inhomogeneities-test, EOFs and Correlation between diferents parameters , Mann-kendall test, correl)

Action: Need quality assesment by expert to take decision then correct automatically by

- a) adding the shift value and document the action
- b) detrend the drift value automatically and document the action.

- **Available software**

USSL- JASL software for hourly data

ETDEVA

QCDAMAR (specific modules for plotting stations and for input/output data)

others

Planification

The JASL software is a very complete aplication in wich most of the procedure describde above are already included but this package are developed under DOS and it will more frienly to have a complete shuite running under windows. For this reason, a complete package has to be developed but taking into account this packages.

- **Methodologies for Quality Assurance**

Inhomogeneity test .standard normal homogeneity test (Alexanderson)

Several tests have been described in the literature, which can be used to detect non-homogeneities in the data series. The SNHT gives the points where an inhomogeneity exists and provides information about the probable break magnitude. But this non-homogeneity could be due to an error or to an anomalous but real behaviour of the variable. For this reason, the series are corrected only by comparison with other series in the same climatic region and supported by historical information about the incidences on the tide gauge.

To perform the SNHT the series are normalised by subtracting the mean and dividing by the standard deviation. The maximum of the following algorithm is the SNHT’s test statistic:

$$T_v = v \cdot (\bar{z}_1)^2 + (n - v) \cdot (\bar{z}_2)^2$$

where \bar{z}_1 is the mean for the series from data point 1 to v, and \bar{z}_2 is the mean from v to the end, n.

Bibliography:

A homogeneity test applied to precipitation data.

Alexanderson, H. (1986)

Journal of Climatology, vol 6, 661-675

Homogeneization of Swedish temperature data. Part 1: Homogeneity test for linear trends

Alexanderson, H. and Moberg, A. (1997)

International Journal of Climatology, vol 17, 25-34

Homogeneization of Swedish temperature data. Part 2: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861

Moberg, A. and Alexanderson, H. (1997)

International Journal of Climatology, vol 17, 35-54

Mann-Kendal Test

According to Sneyers (1975;1992) this non-parametric test can be used to study the increases and decreases in climatic time series.

The Mann-Kendall trend test (Mann, 1945; Kendall, 1975) was used to determine temporal trends in climatic data sets. This non-parametric test has several advantages:

- 1) The data do not need to conform to a pre-specified distribution;
- 2) Missing data are allowed in the data set if they are missing at random; and
- 3) Data values reported as 'less than the detection limit' can be used by assigning them a common value smaller than the smallest measured value in the data set (Gilbert, 1987).

All 'below detection limit' observations were assigned a common value smaller than the smallest detected value in the data set. In the event that multiple detection limits were present in a data set, all non-detect values were assigned the highest detection limit.

If we have the series of rank $x_i=1, \dots, n$, reordered in an increasing way (x_1, x_2, \dots). For each element x_i the number of n_i elements x_j preceding it ($i > j$) is calculated so that $x_i > x_j$. If the series has equals terms, the value n_i is obtained by sum the number of elements $x_i > x_j$ for $i > j$, and the half of the number of elements $x_i = x_j$ for $i > j$.

The test statistic is given by:

$$t = \sum_{i=1}^n n_i$$

The distribution function of t is assumed to be asymptotically Gaussian (null trend test hypothesis) with mean and variance:

$$E(t) = \frac{n(n-1)}{4}$$

$$\text{var}(t) = \frac{n(n-1)(2n+5)}{12}$$

The probability is calculated with a reduced normal law table, α_1 :

$$\alpha_1 = P(|u| > |u(t)|)$$

and

$$u(t) = \frac{t - E(t)}{\sqrt{\text{var}(t)}}$$

Then, if α_0 is the elected significance level,

- $\alpha_1 > \alpha_0 \Rightarrow H_0$ is accepted
- $\alpha_1 < \alpha_0 \Rightarrow H_0$ is rejected \rightarrow Trend:
 - $u(t) > 0 \Rightarrow$ Increasing trend.
 - $u(t) < 0 \Rightarrow$ Decreasing trend.

If a data series shows a significant trend, it is possible to know the moment in which this trend starts to appear by means of a progressive series analysis. For this it is enough to calculate the t value for the first i -terms:

$$t_i = n_1 + n_2 + \dots + n_i$$

The corresponding $u(t_i)$ value is given by:

$$u(t_i) = \frac{t_i - E(t_i)}{\sqrt{\text{var}(t_i)}}$$

and the mean and variance are:

$$E(t_i) = \frac{i(i-1)}{4}$$

$$\text{var}(t_i) = \frac{i(i-1)(2i+5)}{12}$$

After that the point in which the $u(t_i)$ is overstepped, on absolute value, the critic value at 0.05, that is, 1.96.

The same method can be applied to the retrograde series. So, in each term, x_i , the number of elements x_j ($i < j$) with $x_i > x_j$ is calculated as n'_i . Then t'_i is:

$$t'_i = \sum_{i=1}^n n'_i$$

and in the same way:

$$u(t'_i) = \frac{t'_i + \left(\frac{(n-i)(n-i+1)}{4} \right)}{\sqrt{\frac{(n-i)(n-i+1)(2(n-i+1)+5)}{12}}}$$

The intersection between the two lines into the critic values at signification level of 5% shows the beginning of some kind of change. The trend is significant when the curve $u(t_i)$ cross over the significant level. The change is abrupt if the first $u(t_i)$ subseries curve is between ± 1.96 and do

not cross the significance level until the change point (when the second subseries begins); the retrograde series must show an analogue behaviour on the second subseries.

This last application of the Mann-Kendall test, the graphic representation of $u(t)$ along the time axes as for the direct series as the retrograde one, is called sequential version of the Mann-Kendall test.

References:

- Gilbert, R.O. 1987. "Statistical methods for environmental pollution monitoring". Van Nostrand Reinhold, New York, 320 pp.
- Kendall, M.G. 1975. "Rank Correlation Methods", 4th ed. Charles Griffin, London, pp202.
- Mann, H.B. 1945. "Non-parametric tests against trend". *Econometrica*, 13, 245-259.
- Sneyers, R. 1975. "Sur l'analyse statistique des séries d'observations" N° 415 WMO. Ginebra.
- Sneyers, R. 1992. "On the use of statistical analysis for objective determination of climate change". *Meteorol. Zeitschrift*, 1, 247-256.

EOFs

An EOF analysis over a set of tide gauge records (2 or more) is an available method to estimate the trends and to detect possible errors. The stations must be on the same coastal area because in this case the sea response to the forces could be the same along the coast.

Trend determination by EOF can be described by the following steps.

1. The common mode: Commonly, it represents the maximal variance contribution to the sea level series. In this case, the local behaviours due to the special characteristics of each station is expected to remain in another modes.
2. Using the weights, the reconstruction of the common mode at each station can be done. Then,

$$\zeta_i(t) = a_i \cdot P(t)$$

where $P(t)$ is the original series at the station, a_i is the EOF coefficient, and ζ_i is the contribution to the common mode for this station. If the stations are closed enough, the a_i values must be similar.

3. A linear regression for each ζ_i could be calculated. Obviously, if the a_i are similar, the linear trends must be.

In addition, where an EOF contains variance of only 1 station it may indicate errors for these stations. If more than one stations at close proximity are solely represented at an EOF it may indicate regional variability.

Brief summary. Review (novembre 2003):

Control Checks:

QC0: metadata: apply to the header

QC data-checks: apply to the heights values

- Date & time in consecutive order
- Measurement system stability & calibration practices
- Broad range
- Spikes
- Shift and drift time

QC- time series: apply to the means values

- Land stability: shift and drift datum. Need to be implemented
- Inhomogeneities, EOF, Trend-test, Correlation, etc.

QC-Visualisation: data with the corresponding flag in colour.

More of the above quality checks are well known. Some methods are described below:

- **Spikes:**
New Method: fitting of a spline to a moving window of data
*Action: a) **flag** = 5 on values that differ more than N sigmas*
b) Computation of the meteorological residual & application of fitting method to the residual data.

- **Inhomogeneities:**
Objective: find the inhomogeneity in the climate series to detect a probably datum shift.
Gives the break values. Can be apply to the station or to the difference of two stations.
Method:

$$T_v = v \cdot (\bar{z}_1)^2 + (n-v) \cdot (\bar{z}_2)^2$$

The jump is obtained by:

$$\Delta = \sigma (\bar{z}_{2m} - \bar{z}_{1m})$$

- **EOF: Empirical Orthogonal Function.**
Objective: estimate the variances and trends.
Can be used to detect possible errors
Method: calculate the common mode: Commonly, it represents the maximal variance.
Reconstruction of the common mode at each station.
Calculate the linear regression for each ζ_i

- **Mann-Kendall test for trends**
Objective: Study the trend friability. Apply to the original and retrograde series.

Case 1: $\alpha_0=0.2781$, $\alpha_1=0.7219 > \alpha_0$ (not reliable trend)

Case 2: $\alpha_0=1.0000$, $\alpha_1=1.0062 e^{-9} < \alpha_0$ (reliable trend)