ON THE NON-PARAMETRIC CORRECTION OF WAVE FIELDS

Sofia Caires

Meteorological Service of Canada and Royal Netherlands Meteorological Institute

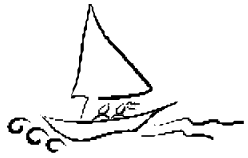
Val Swail

Meteorological Service of Canada

Important/motivational references:

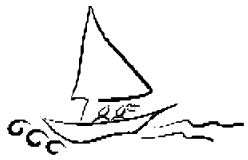
•Swail, V. R. and A. T. Cox, 2000: On the use of NCEP-NCAR reanalysis surface marine wind fields for a long-term North Atlantic wave hindcast. *J. Atmos. Oceanic Technol.*, **17**, 532-545.

•Caires, S. and A. Sterl, 2005: A new non-parametric method to correct model data: Application to significant wave height from the ERA-40 reanalysis. *J. Atmos. Oceanic Tech.* (in press).



Plan of the talk

- Motivation
- Objectives
- Description of the non-parametric approach
- The ERA-40 application
- The present application
- Results
- Conclusions



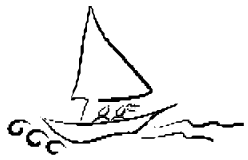
Motivation

Presently there are 2 state-of-the-art global wave reanalysis data sets available:

1. Cox&Swail (2001), computed off-line from the NCEP/NCAR reanalysis (NNR) from 1958 to 1997.
2. The ERA-40, from 1957-2002.

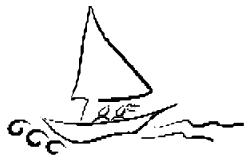
Deficiencies in these wave fields led to

1. The kinematical improvement of the NA NNR wind fields and production of a new wave reanalysis for the NA (Swail&Cox, 2000). expensive
2. The correction of the ERA-40 wave fields using a non-parametric approach (Caires&Sterl, 2005, C-ERA-40). ?



Objectives

- Use a non-parametric method to correct the NNR derived significant wave height (Hs)
- Investigate whether the method is as effective as the kinematic improvement of wind fields



Motivation for the non-parametric correction (npc) of Hs

bias = bias (H_s, swell, ...)

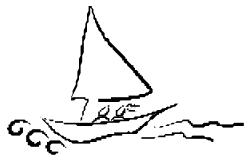
=> no simple parametric correction

hope: bias similar in similar situations

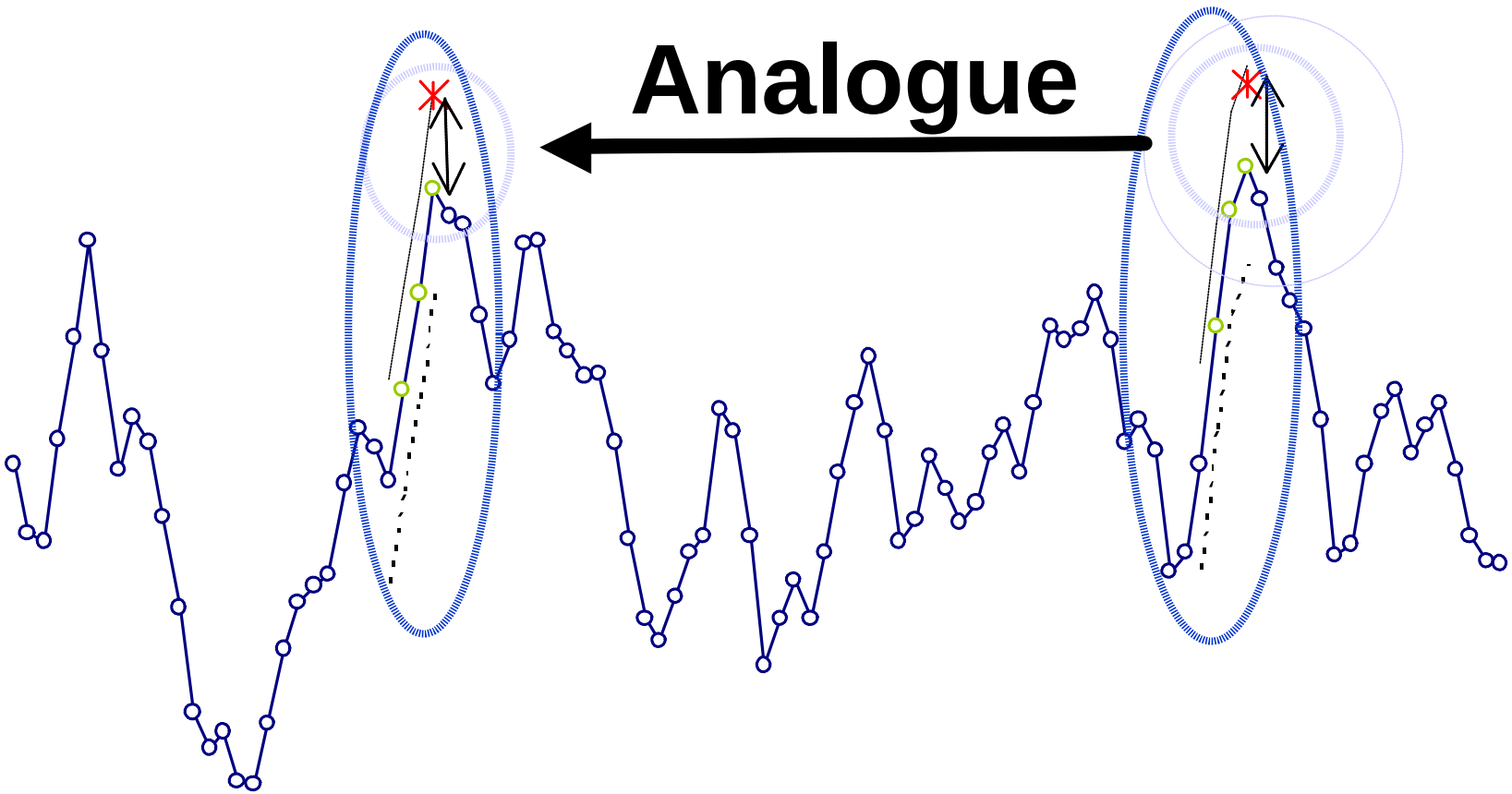
then:

identify “similar” situations (“analogues”)

learn from known biases

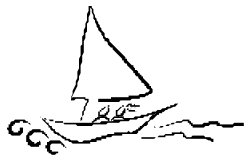


Learning



How to do it

- Divide the data into periods according to inhomogeneities
- “Truth” from TOPEX
- Build learning dataset for each period
- Identify analogues and correct data
- Calculate confidence intervals
- Validate (buoy, Geosat, ERS-2)



Formally

Write $S(u,h) = \{u' \in \Re^m : |u_j - u'_j| < h, j = 1,\ldots,m\}$

for $h>0$, and let $h_n>0$ be a given number depending on the sample size n.

Then, the estimator of the conditional mean of V given U=u, R(u),

is called the *empirical regresssion function* and is defined by

$$R_n(u) = \frac{\sum_{i=1}^{n} V_i 1_{[U_i \in S(u,h_n)]}}{\sum_{i=1}^{n} 1_{[U_i \in S(u,h_n)]}}$$

and the estimator of the conditional distribution function of V given U=u, F(v|u),

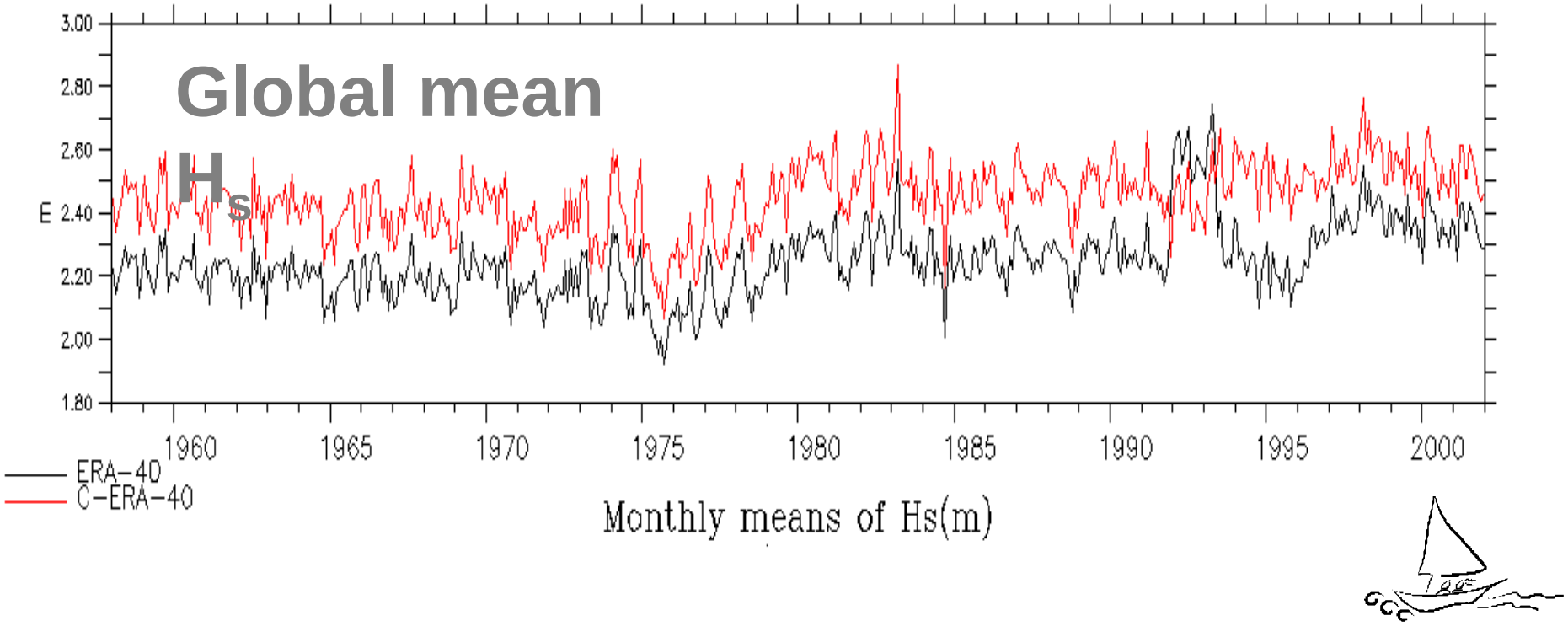
Is called the *empirical conditional distribution function* and is defined by

$$F_n(v \mid u) = \frac{\sum_{i=1}^{n} 1_{[V_i \le v, U_i \in S(u,h_n)]}}{\sum_{i=1}^{n} 1_{[U_i \in S(u,h_n)]}}, \quad v \in \Re$$

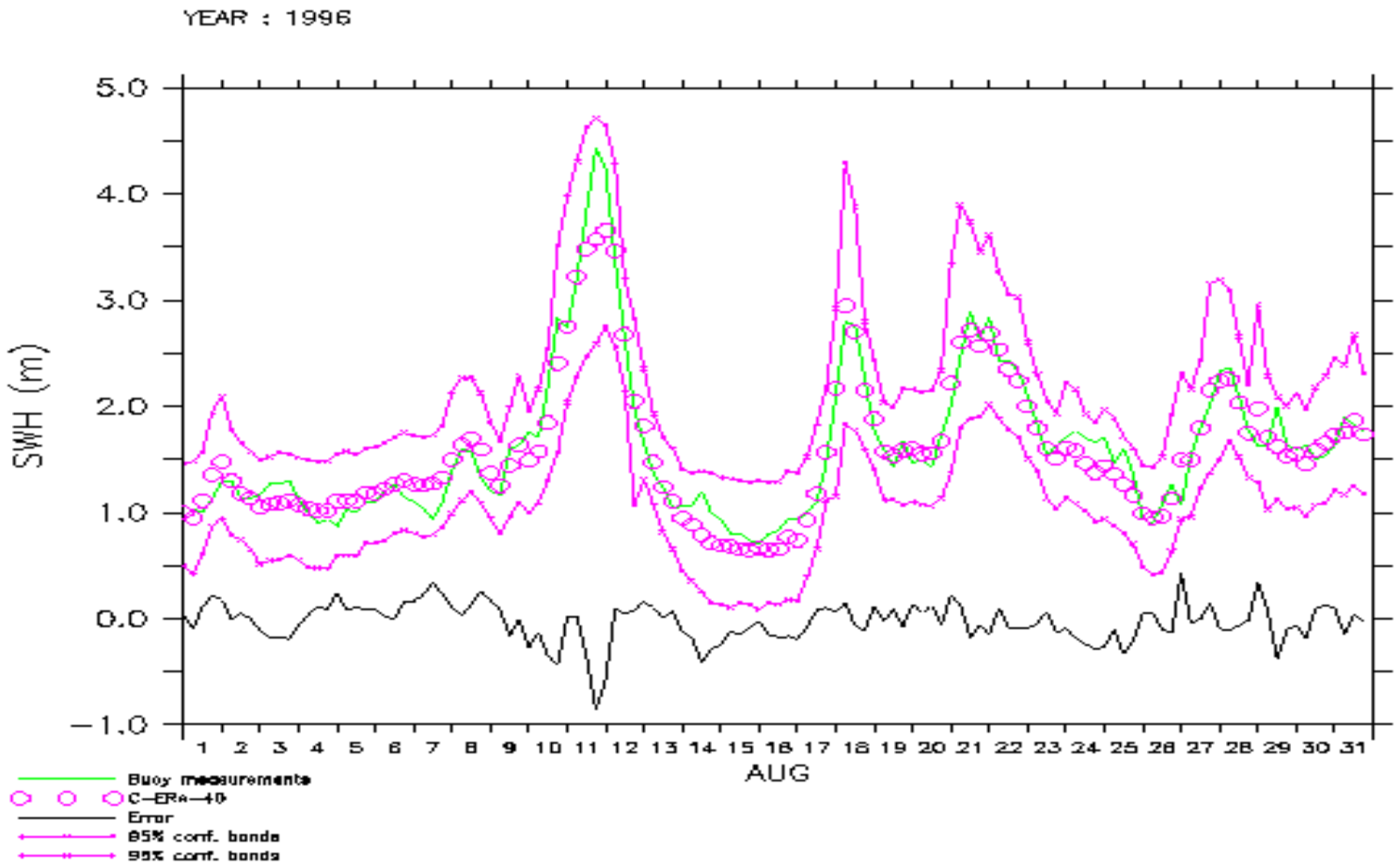
The motivation for using these estimators is that they both converge (as n grows) in some sense and in certain conditions to their theoretical counterparts, R(u) and F(v|u).

Application to the ERA-40 Hs

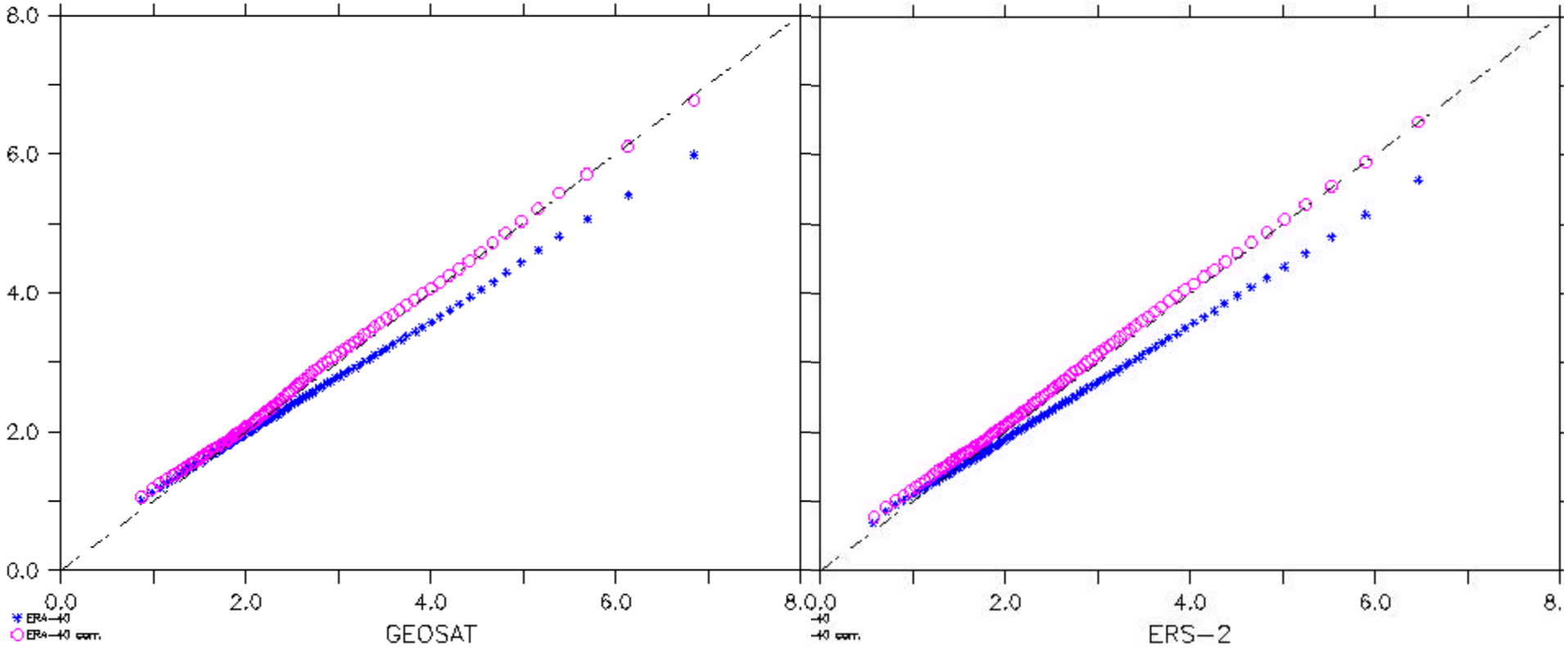
- wave height generally increased (bias ≈ 0)
- no more inhomogeneities



Timeseries


YEAR : 1996
SWH (m)
5.0
4.0
3.0
2.0
1.0
0.0
−1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
AUG
Buoy measurements
C-ERA-40
Error
85% conf. bands
95% conf. bands

1st to 99th percentiles



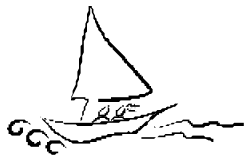
Data from 01/1988 to 12/1988

Data from 06/1995 to 05/1996

Application to the NNR derived Hs

- Two years considered: 1994 and 1997. A learning data set with data from 1997 is used to correct data from 1994 and vice-versa;
- *m=3*: *U* consists of sequences of 3 model values;
- *V* is given by the error between the last value in the sequence and the corresponding Topex measurement, used to estimate the conditional mean and distribution function;
- at each location the learning data set is composed of sequences that are within a 10º circle centred at the location for which the values are being corrected;
- two sequences are considered as analogues if the maximum absolute difference between them is of at most 50 cm, when the sample size of possible analogues is *n*=700;
- Only data from October to March was used to correct October to March and only data from April to September was used to correct April to September. The error characteristics seem to have a seasonal behaviour.

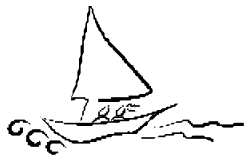
-> C-Cox&Swail



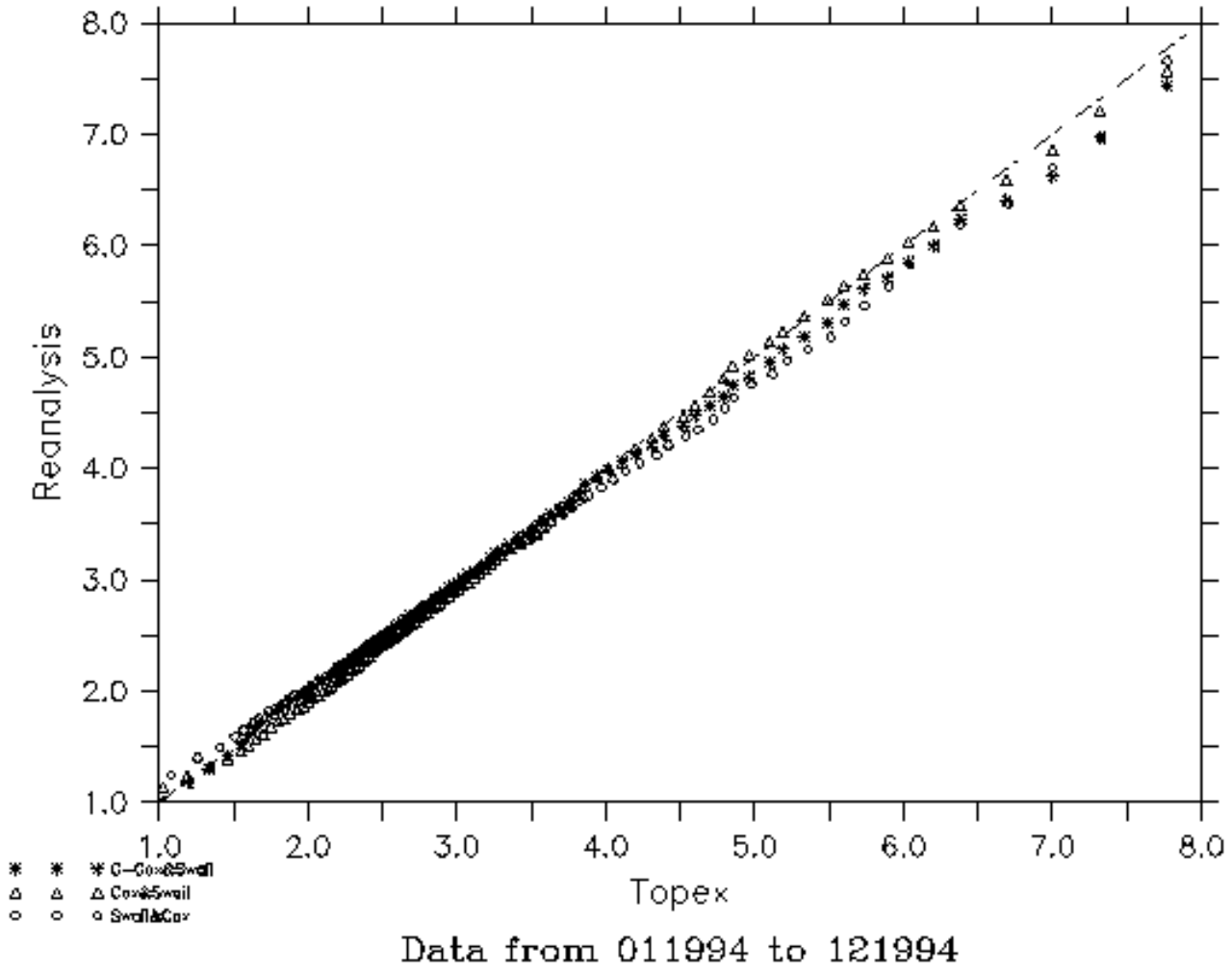
Statistics of significant wave height (m) data for 1994 from different reanalysis products versus Topex measurements in the North Atlantic.

	Bias	**RMSE**	**SI**	**?**
ERA-40	-0.32	0.50	0.16	0.97
C-ERA40	0.00	0.32	0.13	0.97
Cox&Swail	0.04	0.47	0.19	0.94
Swail&Cox	0.03	0.43	0.17	0.95
C-Cox&Swail	0.04	0.44	0.18	0.94

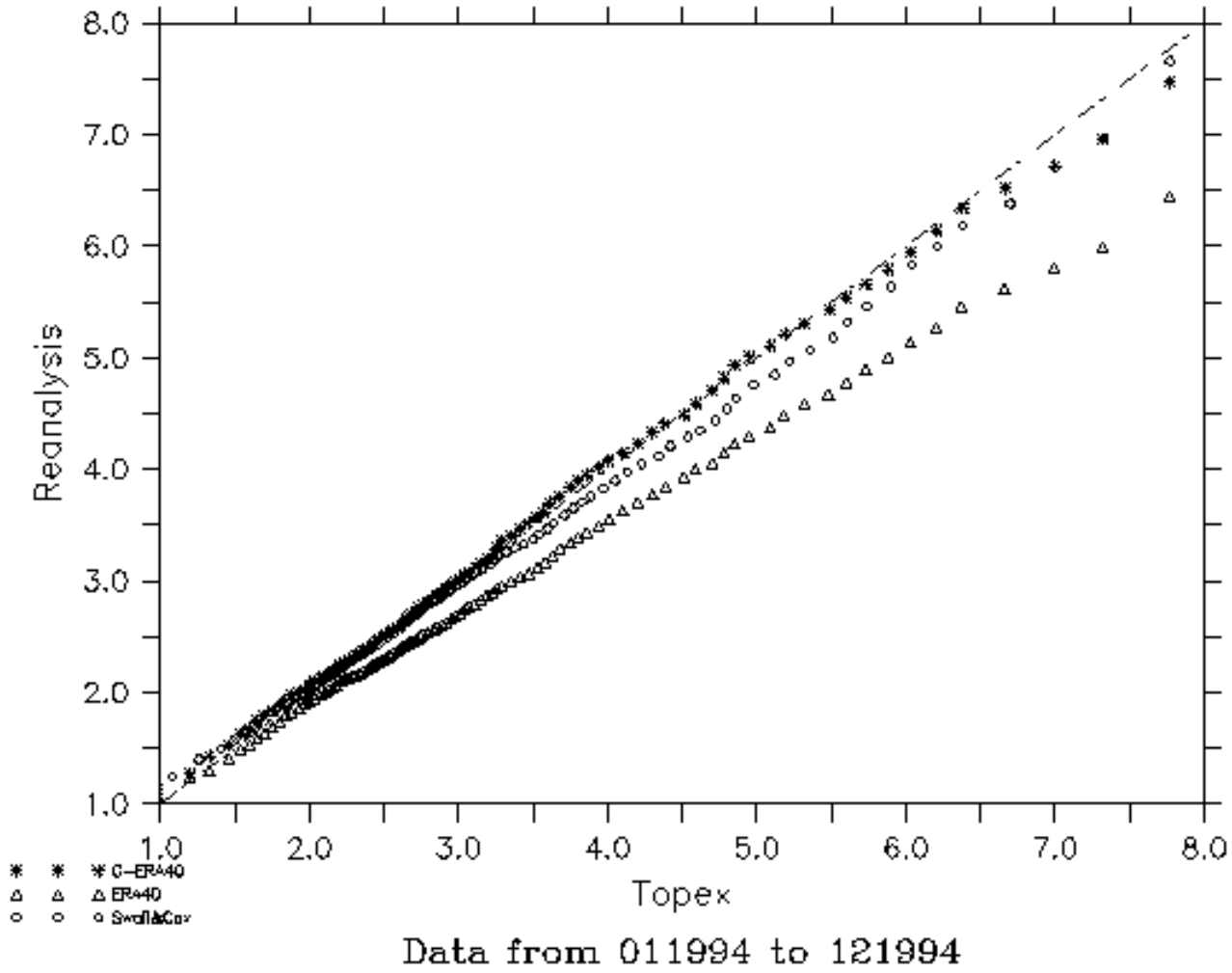
The number of measurements is 49,478 and their average is 2.44 m.



1st to 99th percentiles



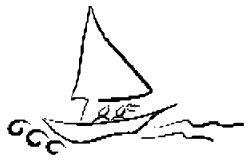
1st to 99th percentiles



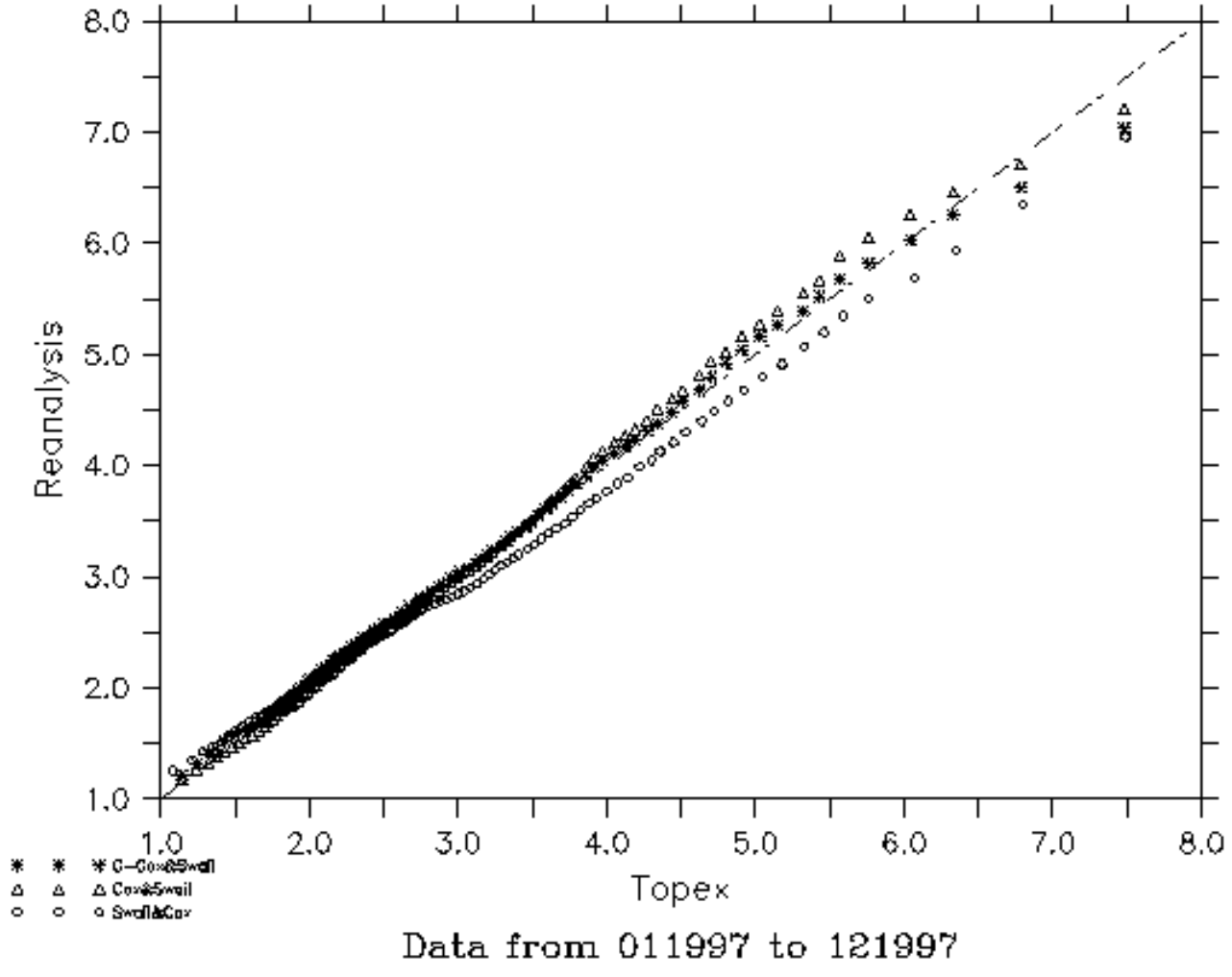
Statistics of significant wave height (m) data for 1997 from different reanalysis products versus Topex measurements in the North Atlantic.

	Bias	**RMSE**	**SI**	**?**
ERA-40	-0.22	0.49	0.18	0.96
C-ERA40	0.00	0.34	0.14	0.97
Cox&Swail	0.00	0.50	0.20	0.93
Swail&Cox	-0.03	0.44	0.18	0.95
C-Cox&Swail	-0.05	0.48	0.19	0.93

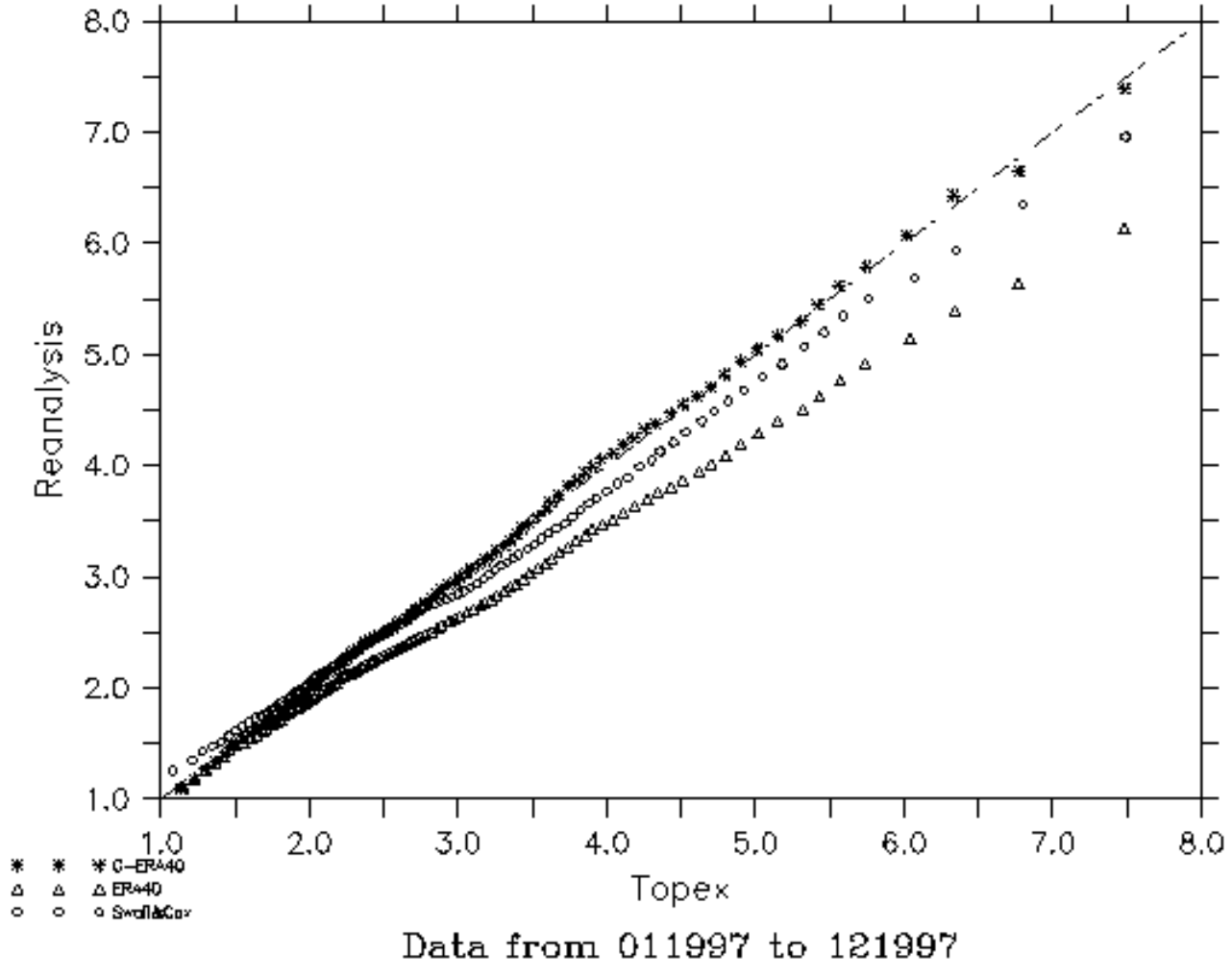
The number of measurements is 48,054 and their average is 2.45 m.



1st to 99th percentiles



1st to 99th percentiles



Conclusions

1. The C-Cox&Swail data set compares better with the Topex data than the Cox&Swail data set. The improvements are, however, small.
2. The Swail&Cox data set compares better with the Topex observations than the C-Cox&Swail data set, although only marginally.
3. The comparisons with buoy data show that the npc has almost no impact in the Cox&Swail data in the NWA buoy locations. The impact in the data for the NP buoy locations is, however, quite visible.
4. The improvements of ERA-40 data obtained by the npc were much more substantial than those obtained for the Cox&Swail data.
5. We suspect that the correction of the Cox&Swail data set was less successful because the correlation between the Cox&Swail data and the observations is smaller than in the case of ERA-40.
6. The npc will not introduce missing storms, remove fake features nor displace storms, and therefore it is important that even if errors are gross the correlation between the data sets is high. This type of errors can, on the other hand, be corrected kinematically.
7. The results of the npc could probably be further improved by extending the learning data set or by adding the wind speed in the conditional setting, but the results will most definitely fall short of the quality of the C-ERA-40 data set.