Met Office

Current and future verification methods

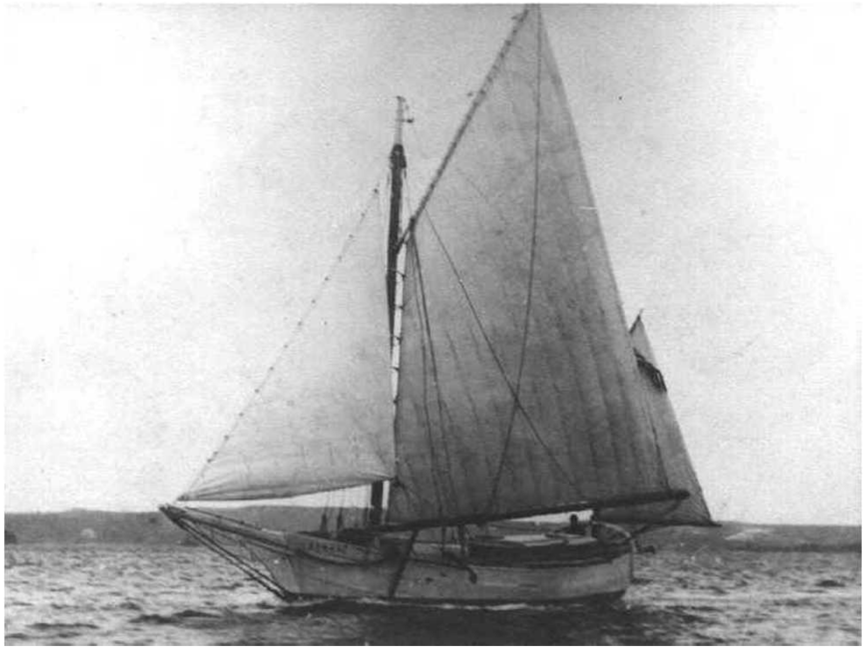
Presentation to ECMWF Wave Forecasting Workshop

Andy Saulter, 25/06/2012

Introduction

'To face the sea is, to be sure, no light matter when the sea is in its grandest mood. You must know the sea, and know that you know it, and not forget that it was made to be sailed over'

– from 'Sailing alone around the world' Joshua Slocum (1st solo circumnavigator – 1895-1898)



Introduction

- This talk will concentrate on verification of operational forecast systems of deterministic forecasts – achieved using monthly, seasonal, annual statistics from regularly changing systems
- Ensemble prediction system verification will be deliberately ignored (a talk in its own right…)
- Operational verification has a key role serving both scientific and downstream user communities and review should ensure that the issues it aims to address are contemporary and methods used are drawn from developments made for long term and case studies
  - Systems monitoring – performance, long term changes and exceptions
  - Uncertainty information for applications by forecasters and downstream users

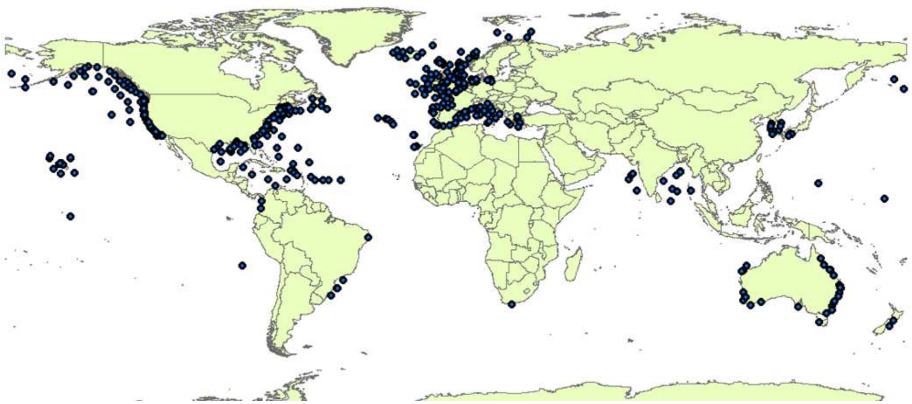
Contents

This presentation covers the following areas

- Introduction
- Changes to the available data
- Reviewing the purpose of operational verification
- What can be addressed?
  - Improving stratification
  - Increasing the parameter space
- Issues
  - Representation errors and scaling
  - Sample independence
  - Are we using a ‘consistent’ truth?
- Summary remarks

Changes to the available data – in-situ buoy data

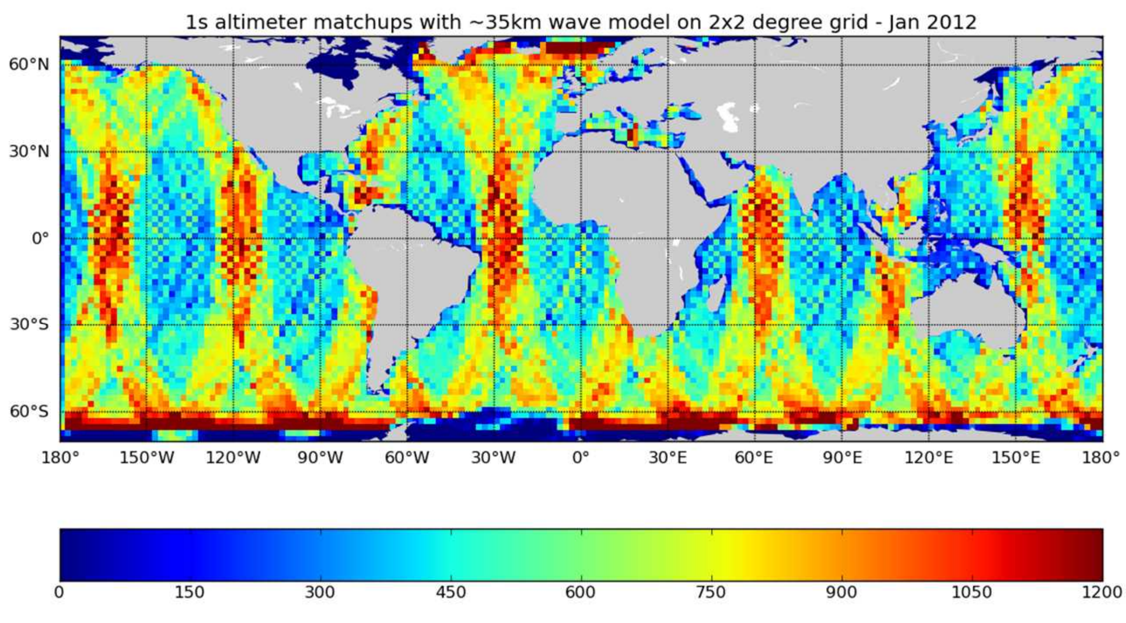
- JCOMM buoy intercomparison first set up in 1995, by 1996:
  - 36 buoys; wind speed and direction, Hs and period
  - 5 Participating centres – global models resolved at 0.5-2.5 degree scales
- Jan 2012 JCOMM buoy intercomparison
  - 399 Buoys and platforms
  - 17 Participating centres – ‘run what you brung’ approach includes 0.25-0.5 degree global models and mesoscale domains



JCOMM buoy intercomparison site list, Jan 2012

Changes to the available data – satellite remote sensing

- Altimeter and ASAR data; significant body of research to demonstrate utility and calibrate against in-situ data since mid 1990s (summaries in GlobWave handbook and Hasselmann et al. 2012)
- Employed in a number of operational centre wave analysis schemes
- GlobWave intercomparison commenced in 2010



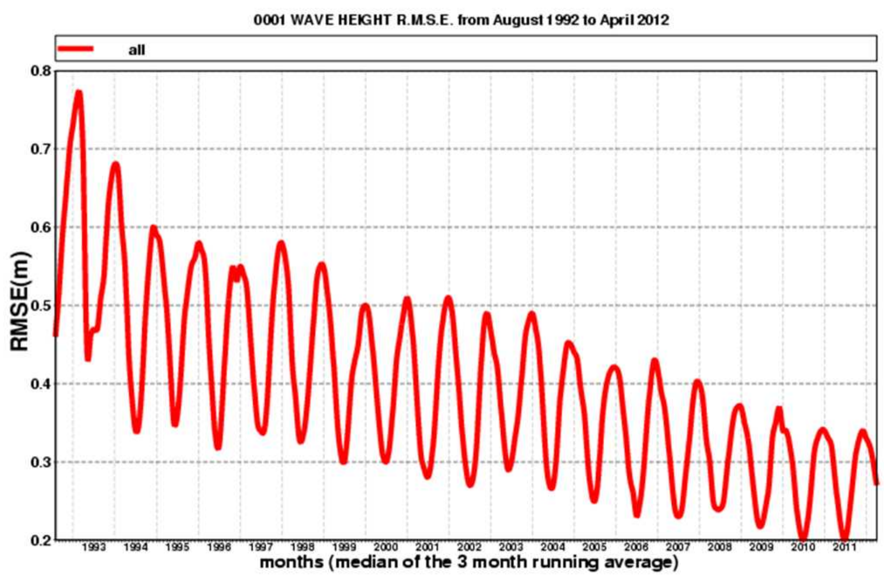
GlobWave WFVS Hs match-ups for altimeter NRT data vs Met Office global wave model

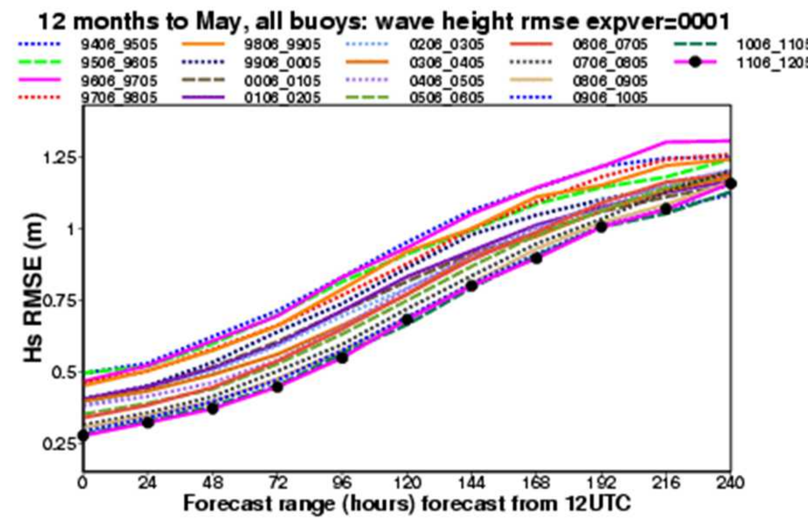
Changes to the available data – new parameters and observations

- Spectral data buoys (approx. 180 available to the JCOMM intercomparison, 100 offshore) – frequency distribution of ‘first five’ components (Jensen et al. 2011)
- ASAR data – 2D spectra for swell (Hasselmann et al., 2012)
- Wind scatterometer (Durrant and Greenslade, 2011)
- Indirect comparison of mean square slope data from altimeter (Ardhuin et al., 2011)
- HF and X-Band radar (Wyatt et al., 2003)
- GPS signals from drifting buoys (Herbers et al., 2011)
- Microseisms data (Ardhuin et al., 2011)

Changes to the available data – model skill

- The fundamental question of ‘is this model generally fit for purpose’ may have been answered reasonably definitively for some time…



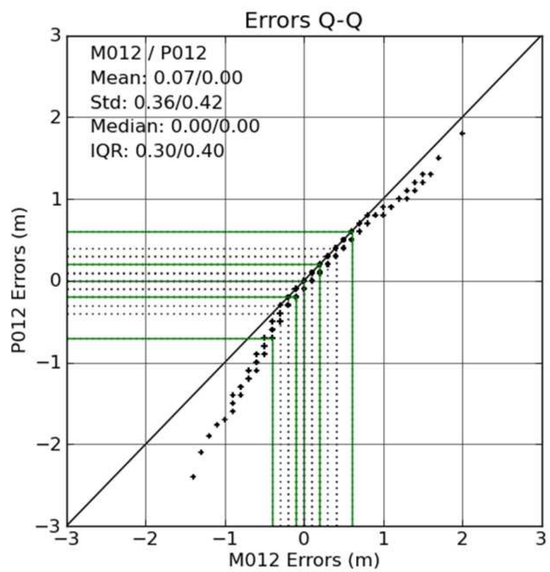


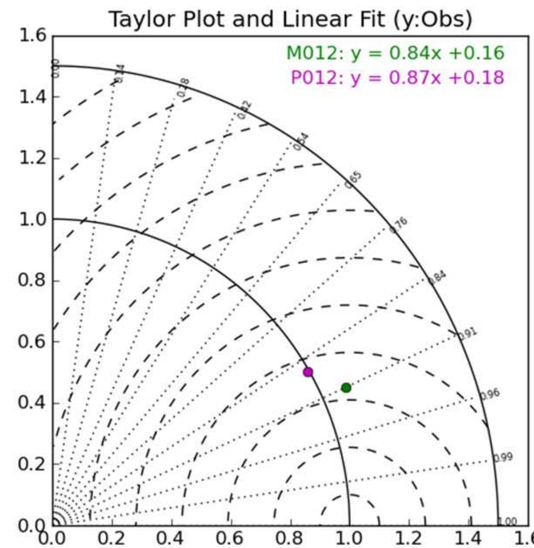
Source: ECMWF, http://www.ecmwf.int/products/forecasts/d/charts/medium/verification/

Changes to the available data – model skill

- The fundamental question of ‘is this model generally fit for purpose’ may have been answered reasonably definitively for some time…

Winter 2012 – UKMO model vs persistence

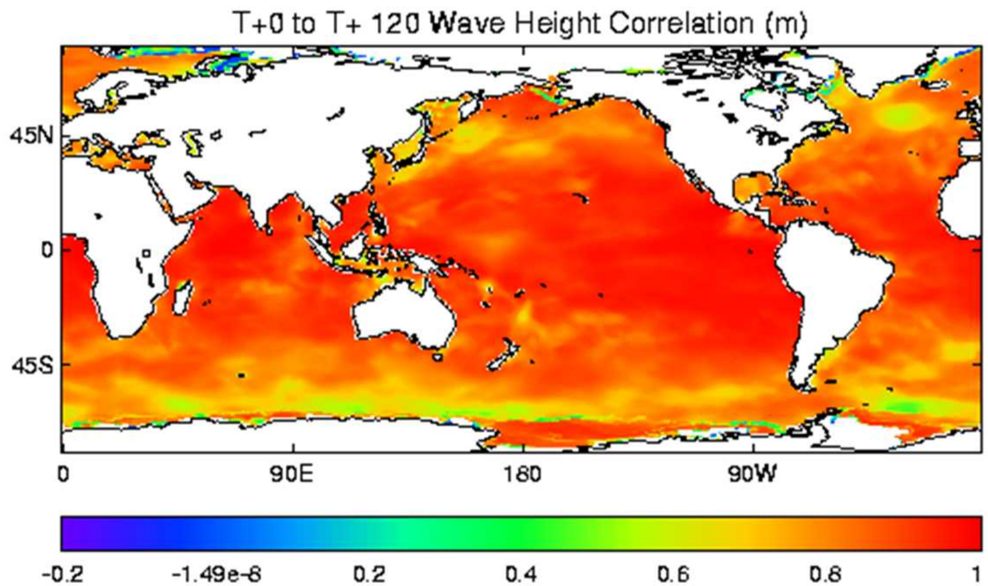




- Implies that present tasks for verification should be focused toward improving details of model performance and providing useful confidence information for any geographic location and wave regime to downstream users

Changes to the available data – model skill

- Options for exchange and comparison of model fields



Hs field correlation between T+0 and T+120 for Met Office global model – Spring 2011

Reviewing purpose - Audience requirements for verification

- Scientist: detail of performance at model scales; effects of adjusting parameterizations and schemes
- Forecaster: detail of performance at observation scales; strategies for intervention and risk communication
- End-user: can I trust the forecast compared to what I experience?
- Different tasks may lead to different requirements in terms of metrics and stratification of data, plus the scales that are verified at…

Reviewing purpose – defining questions to be addressed

- Two elements to metric design:
- Hypothesis statement – parameters to be tested and outcome, e.g. testing improvement of Hs skill, quantifying error between wind forecast and buoy measurements
- Stratification – condition under which the hypothesis is tested, e.g. specific geographic area, moderate to high sea-states, growing wind-seas
- Defining these requirements up front lets us check on the data available, define the best metric(s) to use, get a grip on whether the sample obtained operationally allows stratification
- Observation databases and recently developed model re-analysis/hindcast datasets make the last item practicable

Reviewing purpose – example questions for science audience

- Hypotheses
    - Testing that model performance is improved by system changes
    - Quantifying the errors input to a wave model from the (vector wind) forcing data and other boundary conditions
    - Demonstrating that the overall energy balance in the wave field (as described by Hs) is predicted skilfully and determining any linkages to the input wind errors
    - Demonstrating that energy direction of propagation and spread is skilfully predicted for spectral components of different frequencies.

Reviewing purpose – example questions for science audience

- Stratification requirements (following Tolman et al., 2011) - determine variations in model behaviours for:
  - wind-sea and swell,
  - wave growth under non-aligned winds
  - extreme conditions
  - diminishing winds
  - shallow waters

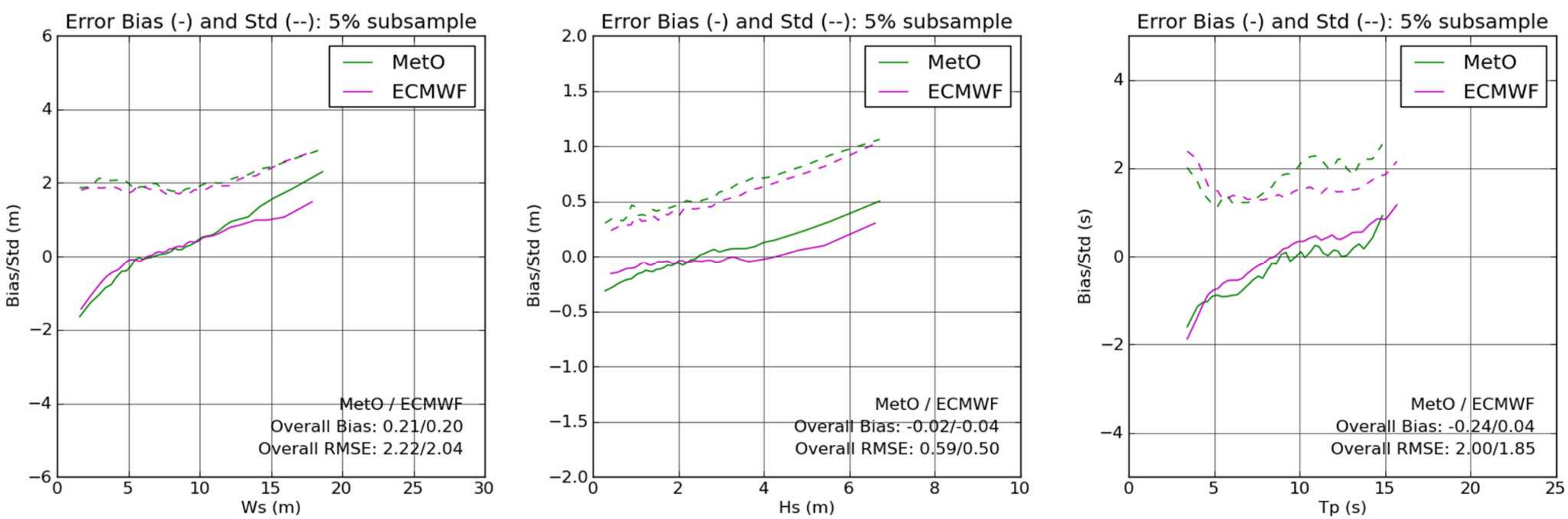
  - stratification against geographic area may also be useful where regions are subject to specific predominant sea types (e.g. short fetch local seas, versus long range oceanic swells)

Reviewing purpose – questions for forecasters

- Hypotheses and stratification will need to relevance to method in which data are applied
- Hypotheses
  - quantify differences between models/lead times
  - quantify distribution of errors
  - are errors consistently predictable across a range of conditions
- Stratification requirements
  - geographic area
  - forecast threshold and lead times (e.g. DNV alpha factors)
  - generating conditions – fetch, wind-sea/swell dominated, wave age

What can be addressed? – improving stratification

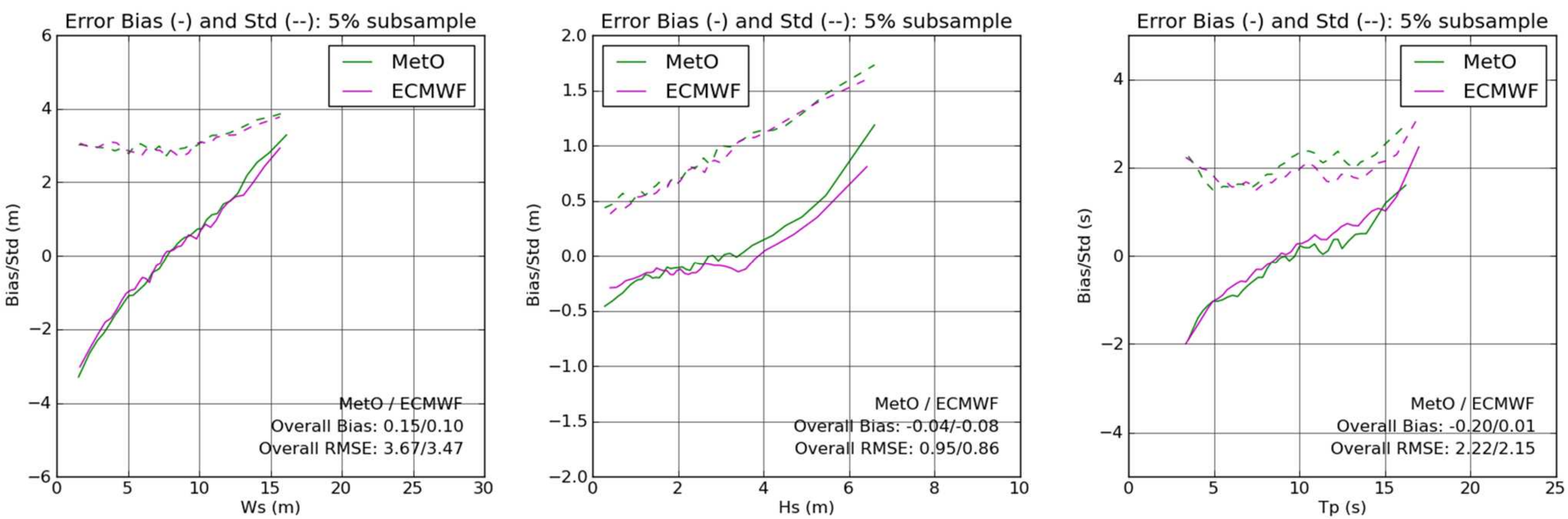
- Increased data volumes allow:
  - greater scope for stratification (conditional, geographic)
  - statistically sensible samples within operational update cycles



Winter 2012 – Met Office and ECMWF comparison to buoy data ~35000 data samples; T+048

What can be addressed? – improving stratification

- Increased data volumes allow:
  - greater scope for stratification (conditional, geographic)
  - statistically sensible samples within operational update cycles

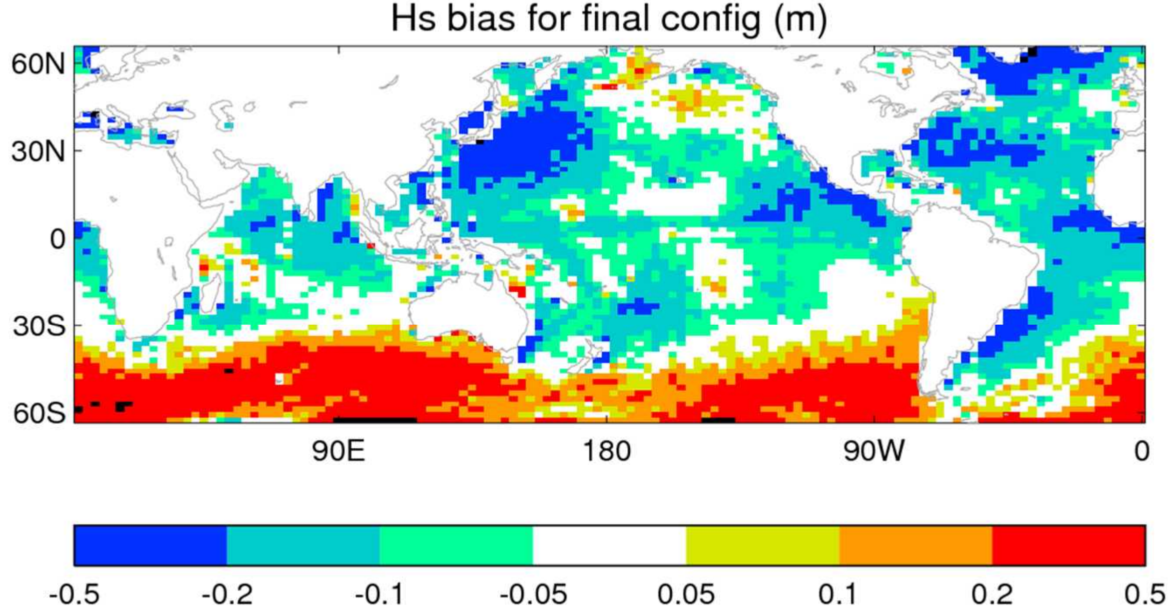


Winter 2012 – Met Office and ECMWF comparison to buoy data ~35000 data samples; T+120

What can be addressed? – improving stratification

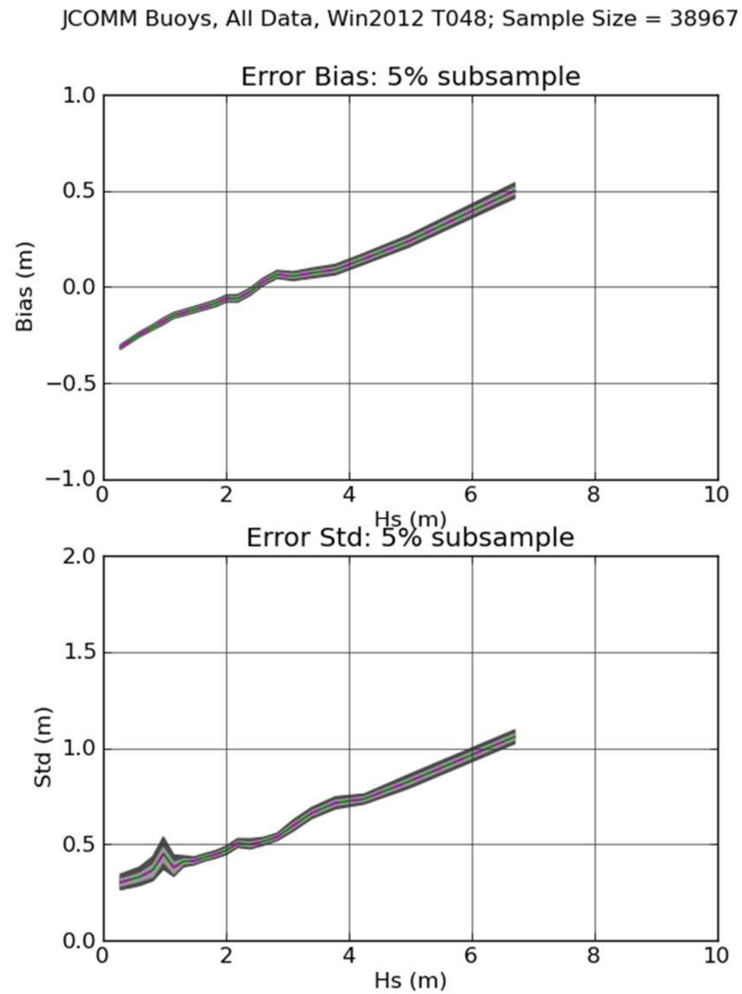
- Increased data volumes allow:
  - greater scope for stratification (conditional, geographic)
  - statistically sensible samples within operational update cycles

1 year (2009) comparison – GlobWave L2P data vs Met Office hindcast; aggregation on 200km grid



Stratification and use of resampling methods

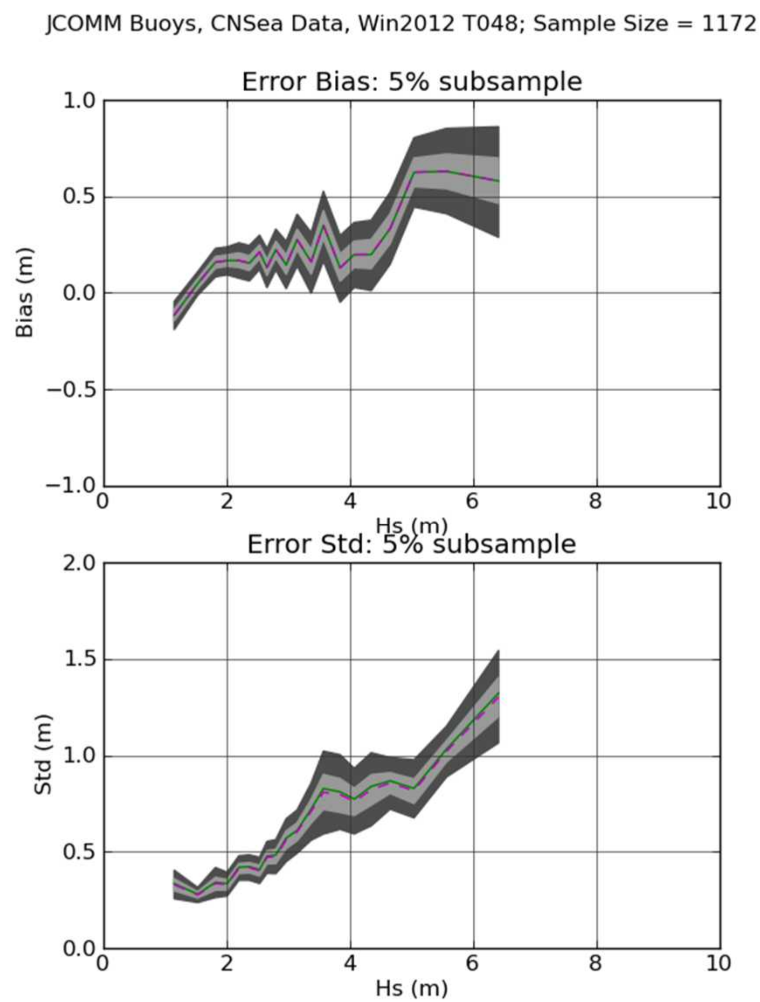
- E.g. bootstrap method to determine variability in statistics using a limited sample



10000 member bootstrap – shaded areas are 50 and 90% confidence bounds

Stratification and use of resampling methods

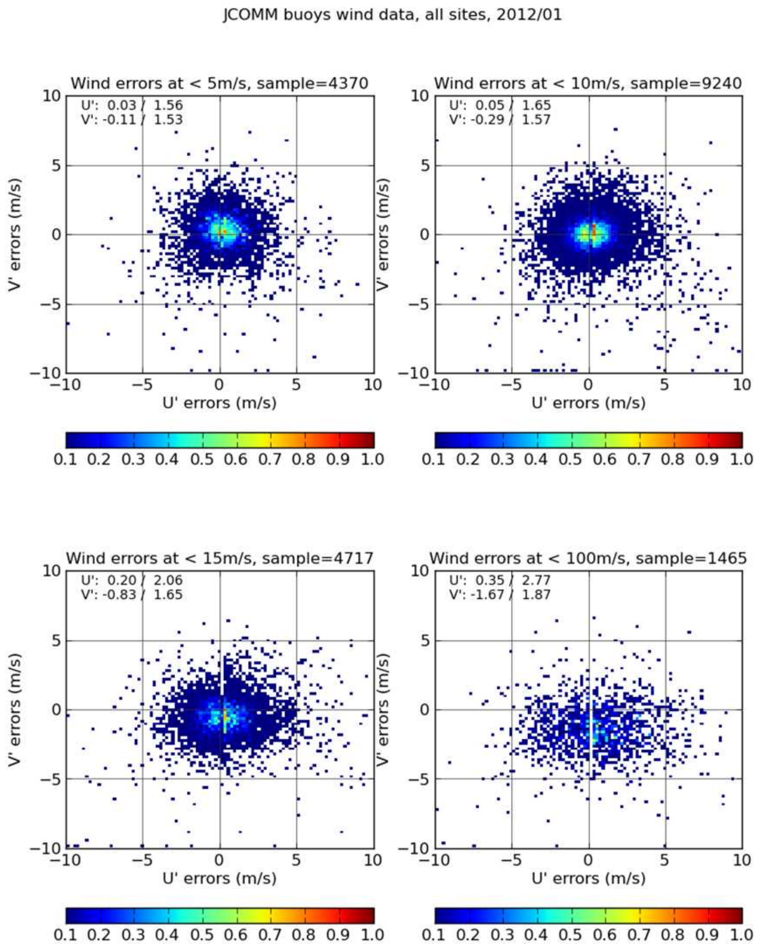
- E.g. bootstrap method to determine variability in statistics using a limited sample



10000 member bootstrap – shaded areas are 50 and 90% confidence bounds

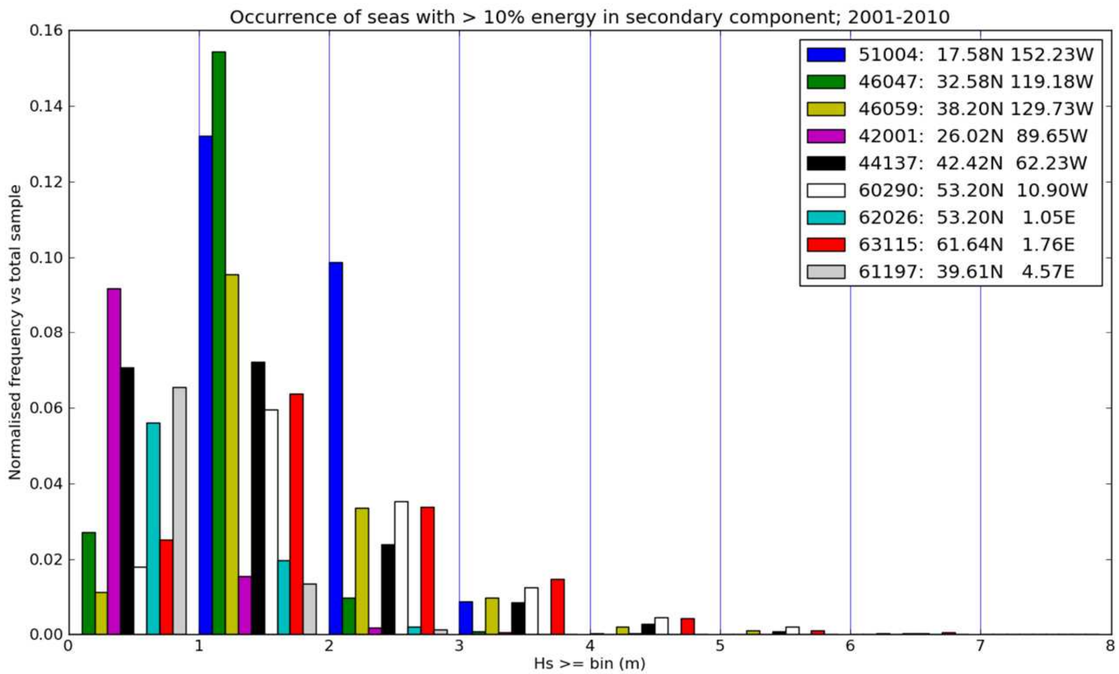
What can be addressed? – increasing parameter space

- Increased sample space needs a larger sample to fill it – difference between what is achievable in operational verification?
- Testing the conditional relationships between primary descriptors of conditions (wind speed, Hs) and errors in secondary descriptive fields (direction, period) possible across high numbers of locations



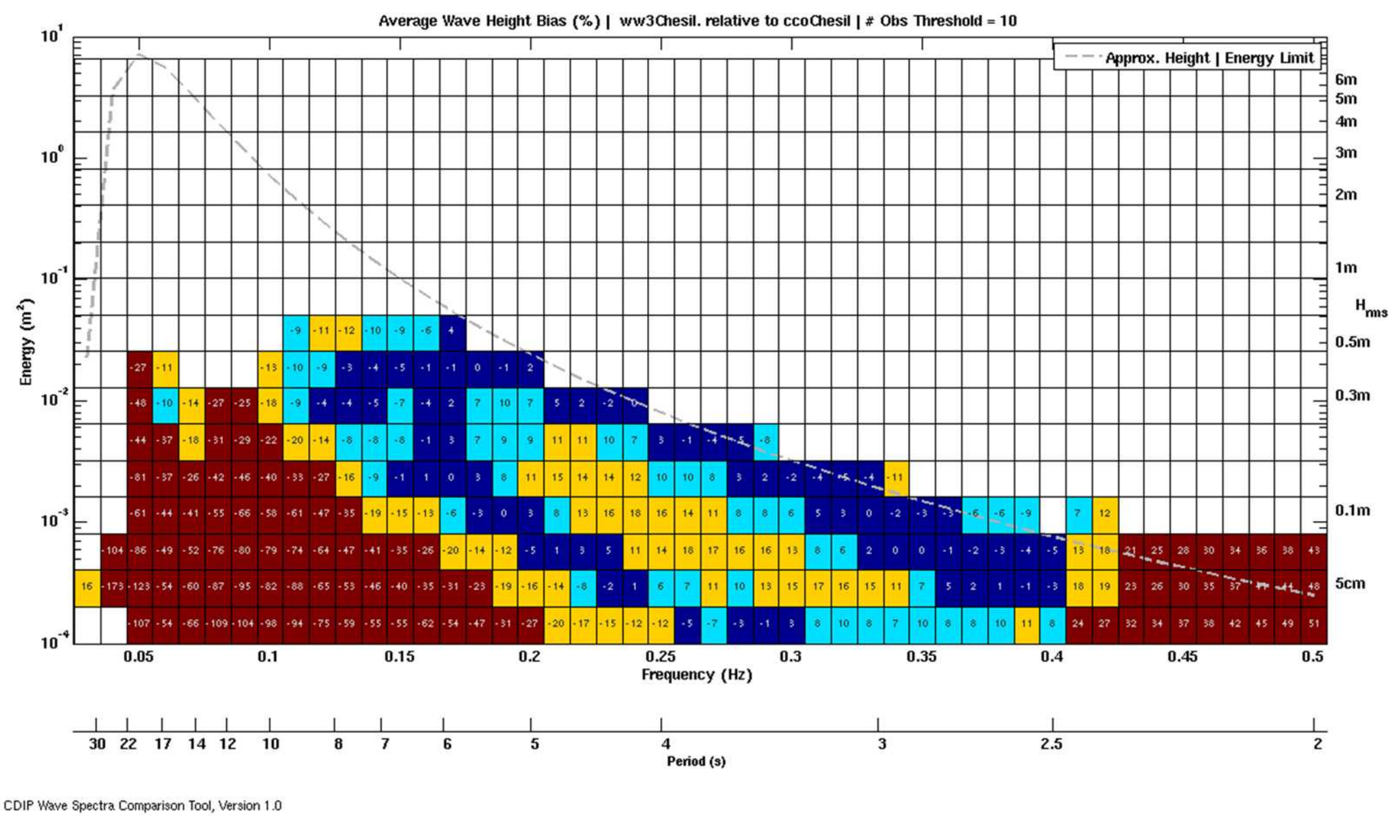
Challenges in interpreting extra sea-state parameters

- ‘Not all 2m seas are born equal’
- Direction and period data sensitive to multi-modal sea-states; use of mean versus peak spectral parameters



What can be addressed? – verification of spectral data

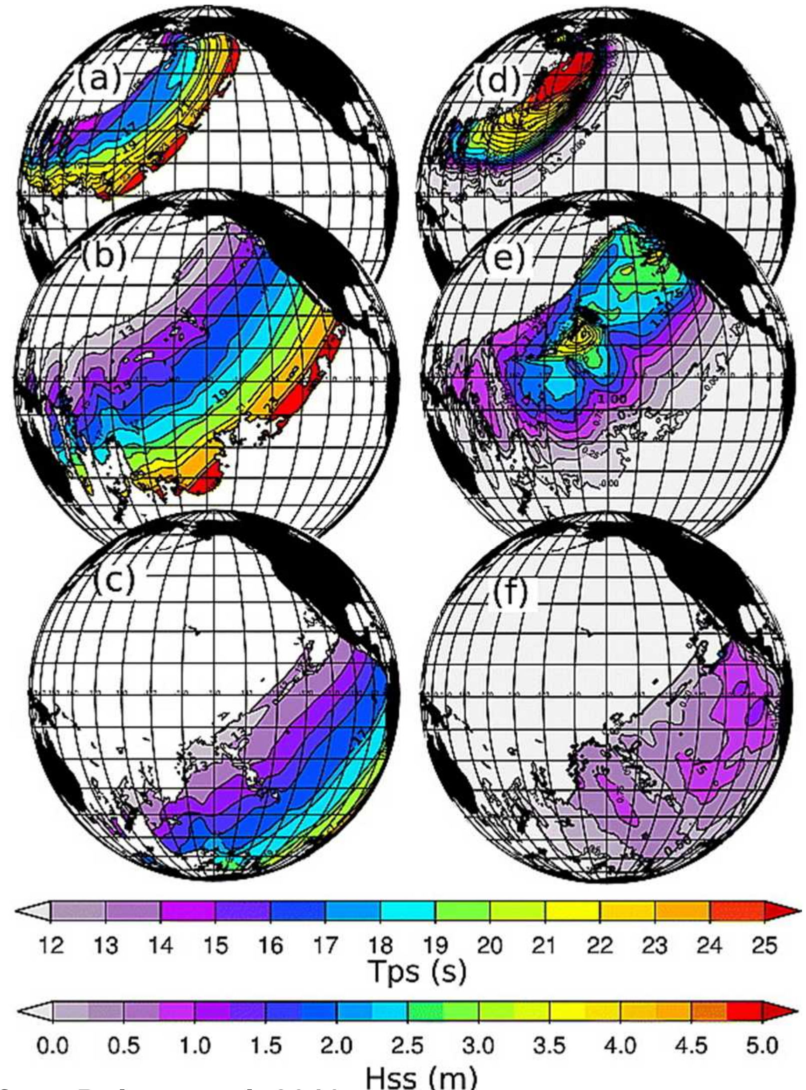
- Methods and tools in existence to verify over various levels of spectral breakdown (Bidlot et al., 2005; Jensen et al., 2011)



Challenges – verification of spectral parameters

Met Office

- Methods and tools in existence to verify over various levels of spectral breakdown (Hanson et al., 2009; Li and Holt, 2009; Ardhuin et al., 2011)
- Long period swell tracking feasible in both model and ASAR data (e.g. Delpey et al. 2010)
- Need to test whether coarser breakdown will answer questions and yield suitable data samples
- Need agreement of methodology



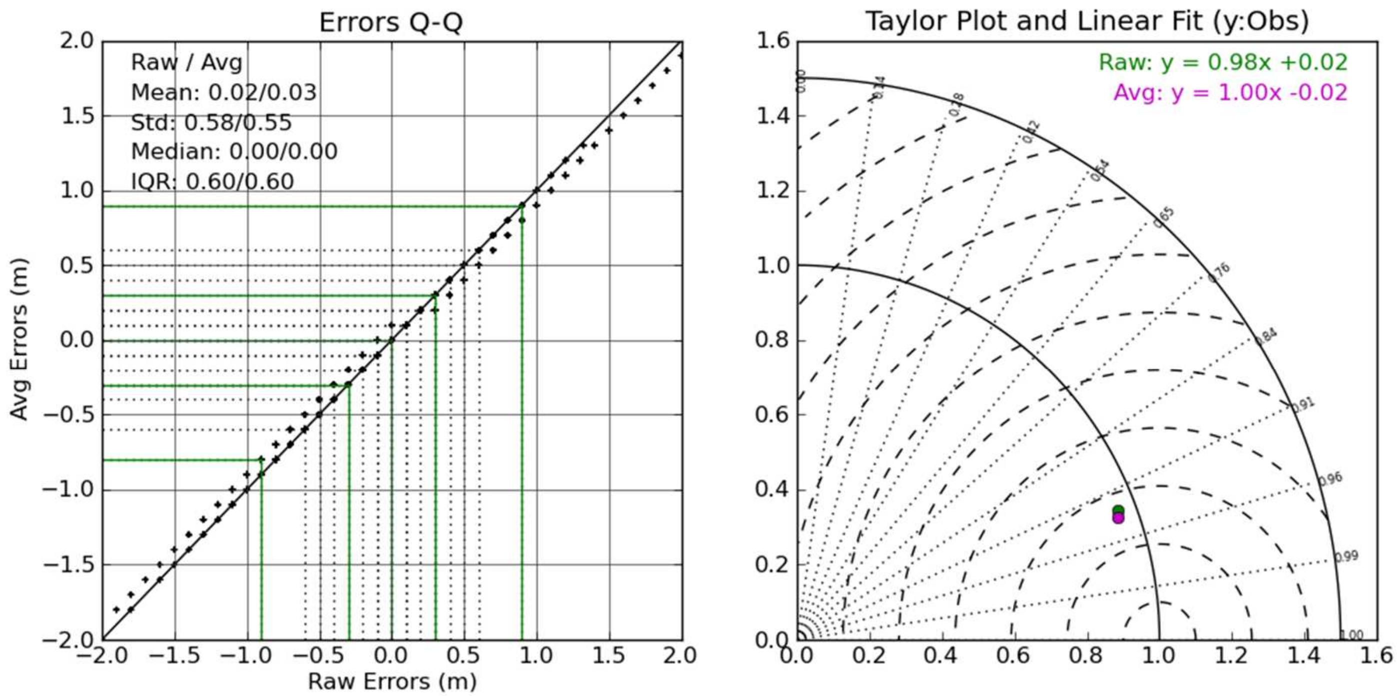
Swell tracking – figure from Delpey et al. 2010

Issues – representation errors and scaling

- Contribution to errors due to representation differences between model scale and period/area sampled in observation
- Super-observing to model scales is used to mitigate representation contribution to overall model – observation errors
- Useful for a scientist assessing an individual model
- For intercomparison of models at different scales, forecaster interpretation and user performance measures using a standard observed scale is possibly more consistent with application?

Issues – representation errors and scaling

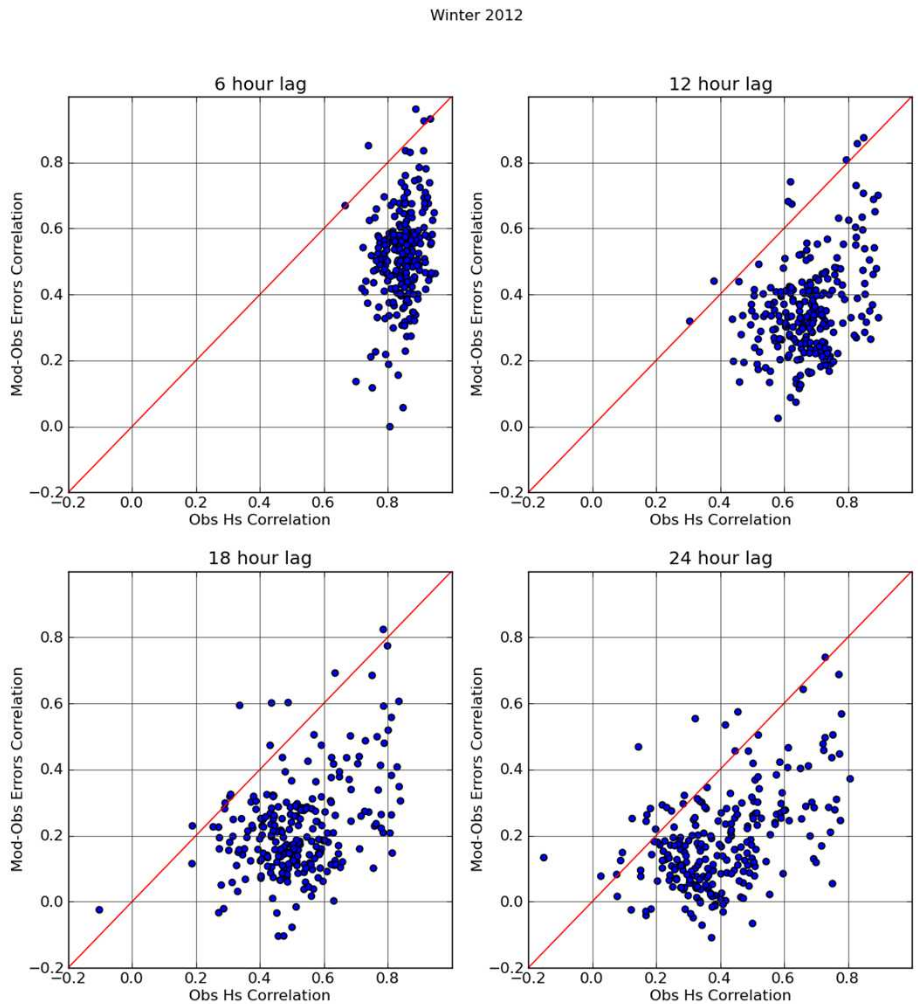
- Besides…



Winter 2012 – JCOMM buoy data, raw and super-observed averages versus Met Office Hs data at T+48

Issues –independence in the observed datasets?

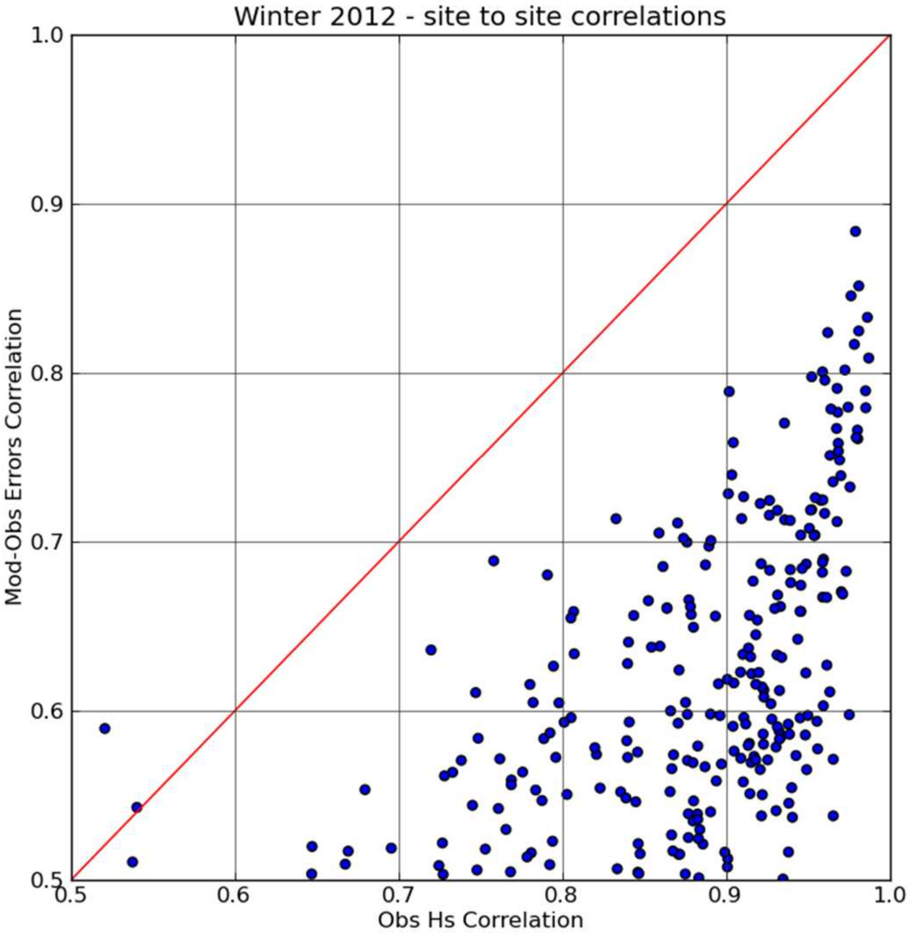
- E.g. time and site correlation in buoy data



Hs observation correlation vs error correlation at different lag times for buoys in the JCOMM intercomparison

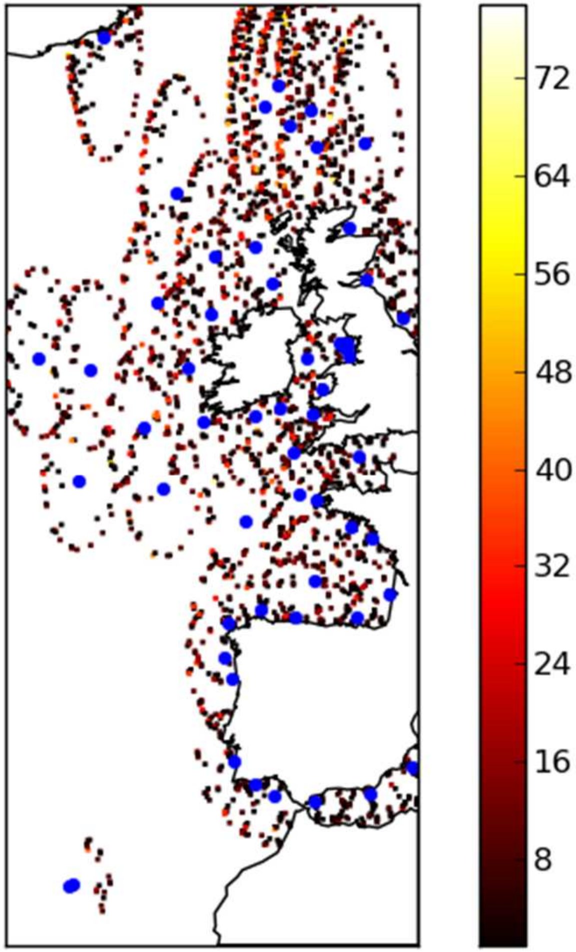
Issues –independence in the observed datasets?

- E.g. time and site correlation in buoy data



Issues - is there a consistent observed truth?

- Ideal for downstream forecasters and end-users – consistent verification data for various operating regions worldwide
- Examples of inconsistencies in observation programmes (e.g. Queffelou, 2006; Durrant et al., 2009; JCOMM PP-WET project)
- Assessment of observing errors (e.g. through triple colocation at global scales as for Janssen et al., 2006)
- Differences may be introduced due to changes in wave climate/generating processes and mix of sampling methods. E.g. North Sea vs Indian Ocean?
- Limits to parameters and metrics based on most common observation types and lowest common denominator sampling period



200km match-up of 1 year unique altimeter passes to NE Atlantic wave buoy network

Summary remarks

- Enhancements in the availability of observations of different wave parameters, model skill and tools to exchange and analyze wave verification data suggest that present operational verification systems could be updated to address questions beyond skill of the Hs forecast
- A principal step in redefining the systems involves clarifying the hypotheses that can be tested within the verification, including stratification requirements
- The need to make outputs applicable to downstream user groups should be considered alongside scientific interests – an issue raised is the approach to representation errors
- Some upfront analyses may be required to deal with sampling issues raised by use of stratification, obtaining independent matchup data, and to understand consistency within a growing observation network
- A key area for development relates to use of spectral or partitioned wave data