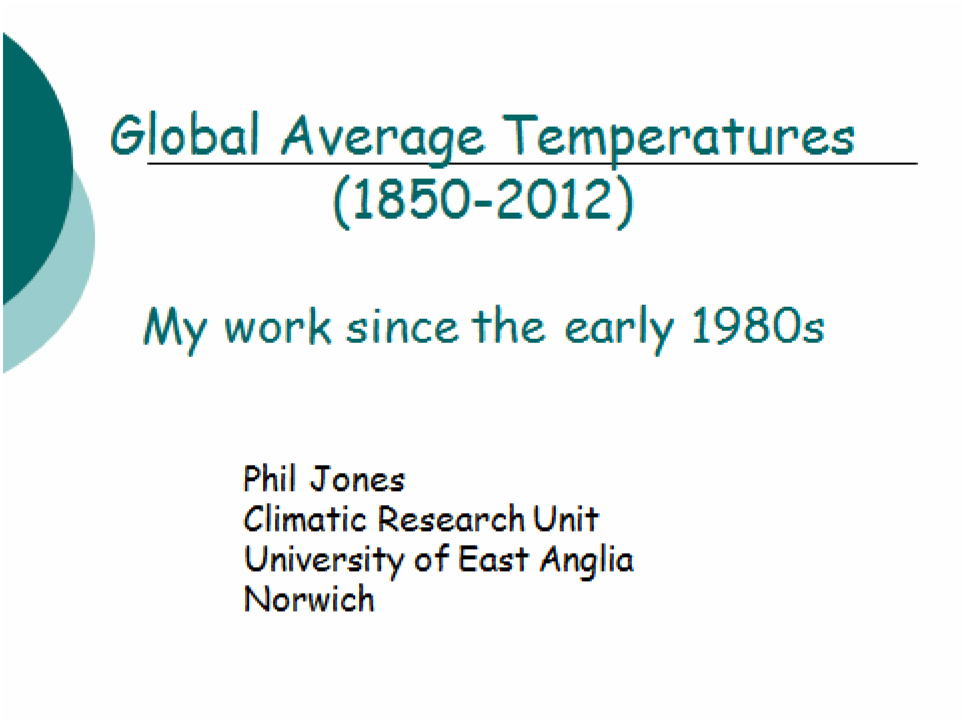
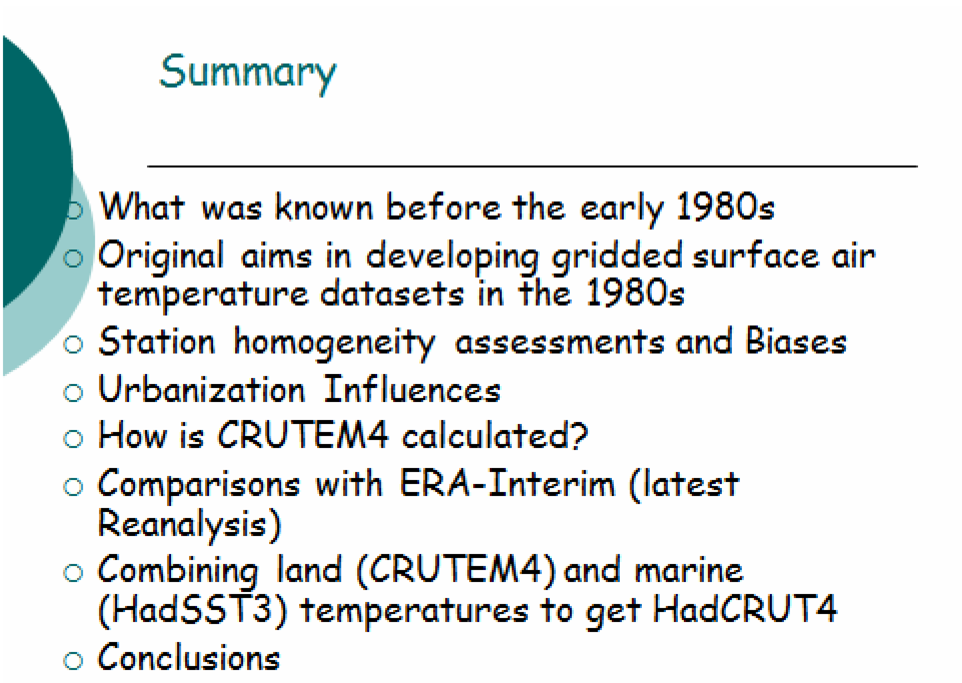
Honourable Rector,

Esteemed president

Good morning everyone



The title of my talk is Global Average Temperatures (1850-2012) and I will emphasize my work and involvement in this subject since about 1980.



The summary of my talk is as follows:

Who had undertaken work on this subject before 1980 and what were there conclusions?

What were my original aims in developing a gridded surface air temperature dataset in the 1980s?

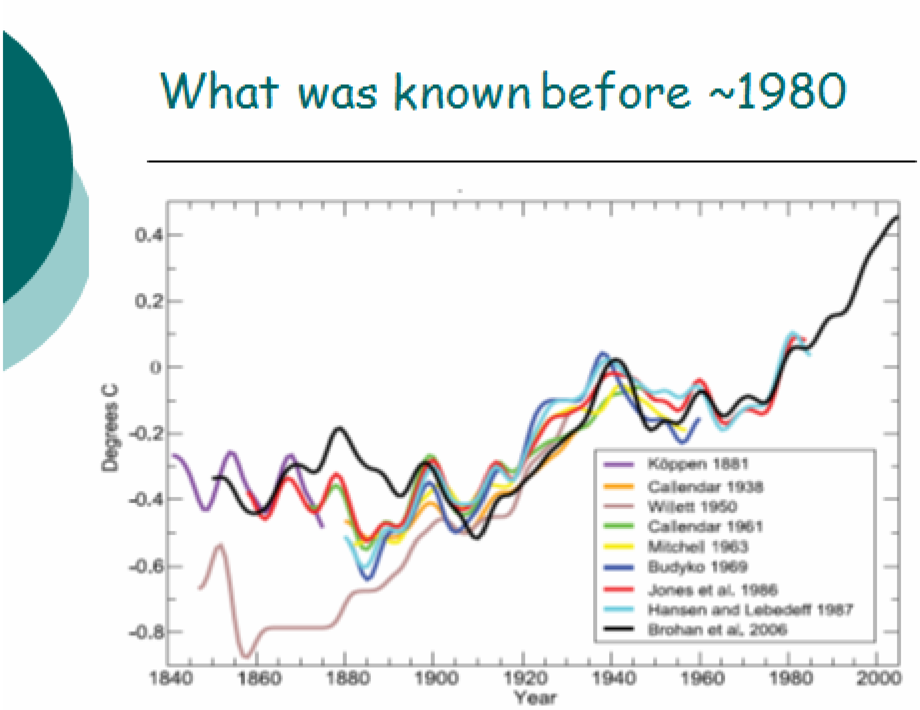
Why do we need to make some modifications to historically recorded temperatures – which we refer to homogeneity adjustments? Homogeneity is the word used by climatologists for the consistency of single station time series, so we can compare all years and months with each other throughout the series. Biases tend to be more important than individual station homogeneity. This is because biases can affect all the data in a similar way. These include the problem of taking sea-surface temperatures measurements and the possible impacts of changes in exposure of land-based temperatures and of urbanization effects around some locations.

How is the latest CRUTEM4 dataset for the Earth's land areas calculated?

Comparisons of this new land dataset with the latest Reanalyses developed by the European Centre for Medium-Range Weather Forecasts (ERA-Interim)

CRUTEM4 covers just the land, so to get a more complete global picture we need to add in sea-surface temperatures (SSTs). This we do with the Met Office Hadley Centre SST dataset called HadSST3. The combined version is referred to as HadCRUT4.

And finally conclusions, with some thoughts on possible improvements.

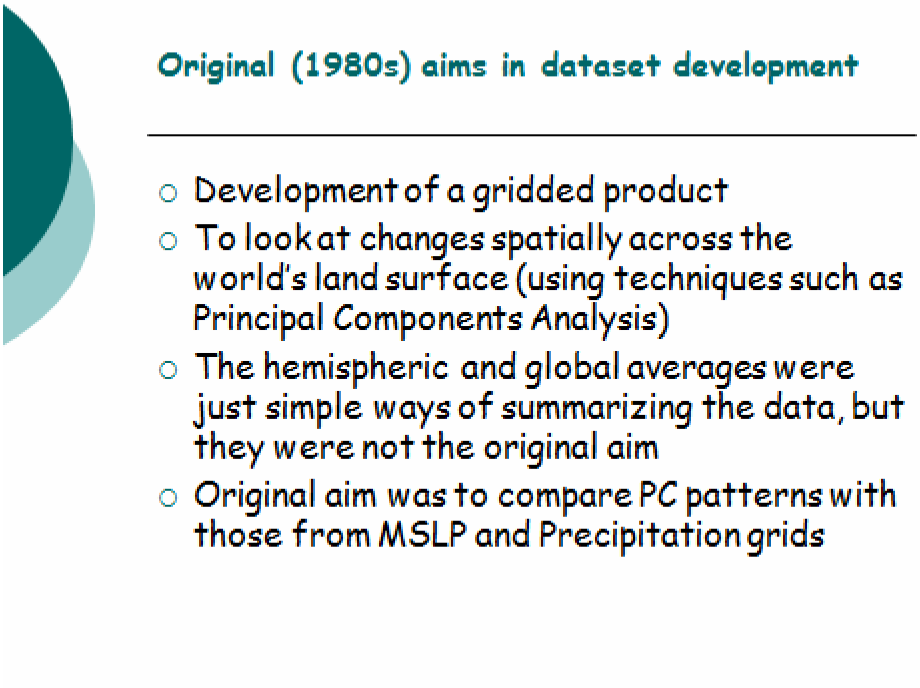


The first person to produce a land air temperature record (for the Northern Hemisphere) was Wladimir Koeppen in the 1870s. Later work was undertaken by others, including Guy Stewart Callendar (in the 1930s and the 1960s) also by Murray Mitchell in the 1960s and Mikhail Budyko in 1969. More recently there are 3 main groups: CRU/UEA, GISS/NASA (Jim Hansen) and NCDC/NOAA.

The first two of these started in the early 1980s, with NCDC beginning in the late-1980s. Series are also produced in Russia, Japan and China using surface temperature data.

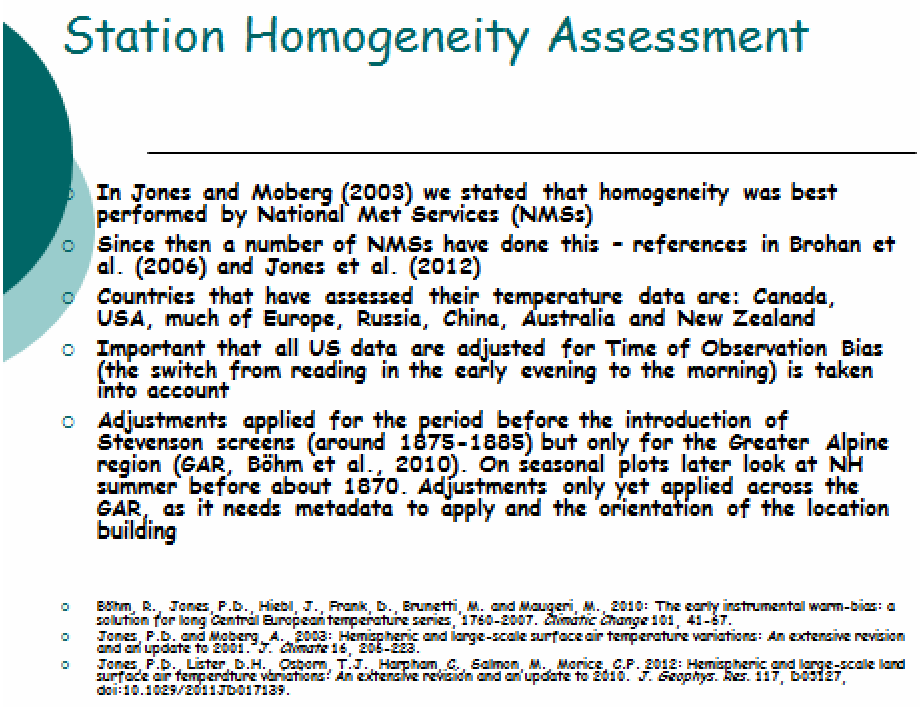
This plot produced by the IPCC Report in 2007 shows that once converted to a modern base period of 1961-90 the early records all generally compare very well, considering that all the analyses before 1980 were effectively done by hand, eventually using adding machines, calculators and the first computers. Much of the basic temperature data comes from World Weather Records (WWR, volumes which began during the 1920s, and which were first digitised in the 1970s). Some European series extend back to the 1700s.

This slide also indicates that when I started (~1980) the temperature of the world had been cooling slightly from the 1940s, but the rise from about 1910-40 was clearly captured by many of the earlier workers. Even Koeppen was quite good, but the apparent 10-11 year cycle he noted meant he erroneously suggested that the variability was due to sunspots. This was based on only 2.5 solar cycles, but it led to numerous papers where scientists kept on looking for solar influences. In the 1970s one scientist (Barrie Pittock) showed that all these papers (and there were hundreds) often used selected periods and those that didn't failed when more data were added. He concluded that you needed 7 solar cycles to be sure you had found reliable relationships!   This is not to say that the Sun isn't unimportant, but you need to base relationships on long series. The only paper that Pittock found that stood up, was one relating the area of the US affected by drought.



I will refer to the temperature dataset as a gridded product. The reason for developing this was to interpolate to a regular grid to avoid having to use the irregularly-located station data. Interpolation produces more complete series, as missing data from one location can potentially be interpolated from neighbouring locations.

Our original aims of developing the gridded product was to look at changes spatially across the world. We already had a similar dataset for sea-level pressure and we wanted to relate this to temperature and to consider techniques such as Principal Components Analysis and relate the time-series scores from the pressure and the temperature data. Producing hemispheric and global averages was just a convenient way of summarizing the data. We did some comparisons in the first paper in 1982 (as in the last slide) but this was not the original aim. Also as stated, when we started the world had been in a cooling phase since the 1940s.



With the first paper in 1982, we didn't take any consideration of the long-term homogeneity of the station temperature series into account. There are lots of reasons why temperature series are likely to be non-homogeneous. These include moves of the station location, changes in measurement procedures (in the times the measurements were taken and how the daily and monthly averages were calculated), changes in instruments and possible effects of recent automation of the measurements (including the switch from mercury-in-glass thermometers to platinum-resistance sensors, particularly the different response times) and changes in the environment around the site (e.g. possible urbanization effects).

We first assessed homogeneity in a series of papers in 1986 and published all the adjustments we made in a series of reports published by the US Dept of Energy (who have funded much of the work over the years). We showed at that time that the overall effect of these station homogeneity adjustments was near zero, as the changes occurred at different sites at different times. We didn't do any more work on this, but in the 2003 update we began using adjusted data developed by some National Met Services (NMSs) and by some fellow scientists. We suggested then that more NMSs work on their national data and many have (much of Europe, Canada, USA, Russia, Australia and NZ). More are doing this but have not completed the work or they don't release the adjusted series. We also suggested that more countries digitize their series and make more of their data available. This is now beginning to happen as well, but some countries are still very reluctant to release any data (India being the largest country that releases very little – even to its own scientists). Some countries still consider that some of their data has some economic value – including my own. Brazil recently released over 300 series of daily temperature data from 1961 to the present.

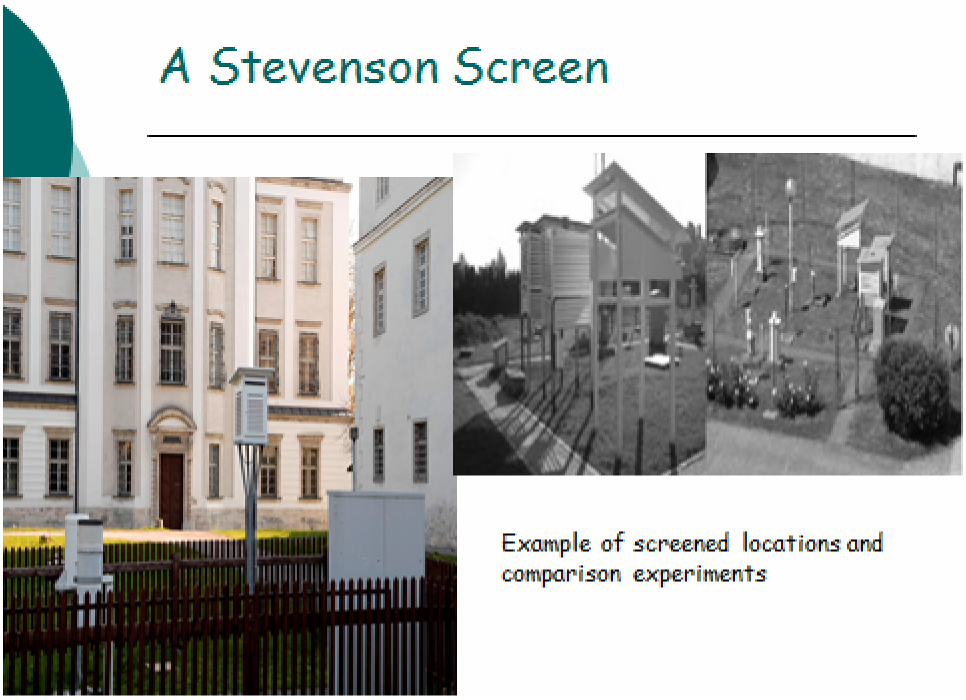
We noted two important aspects that were more pervasive than the other myriad of homogeneity problems. Climatologists refer to these as ‘Biases’ as similar changes are likely to occur at all locations, so simply getting more data doesn’t lead to cancellation of the problems.

The first of these is the Time of Observation Bias (TOB) in the USA which results from observers changing the time of reading from the late afternoon to the early morning. This only affects the USA, but it took place from the 1970s to the 1980s at most locations. This change introduces a cooling bias into the series, which is largest in the interior of the country. This has been extensively adjusted for. The effect of adjustment for this bias is to cool measured temperatures before the change relative to the latter period.

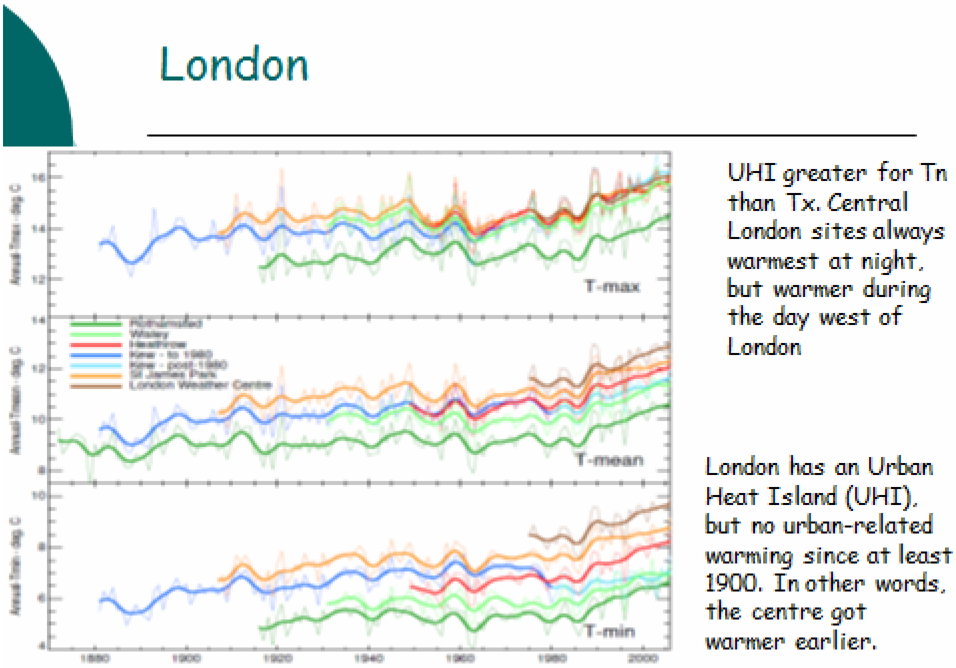
The second major problem affects the measurements before the development of Stevenson-type screens around 1880. Before this instruments were placed on north wall locations in the Northern Hemisphere, but in summer months these instruments can be affected by direct sunlight in the early morning and late evening. This problem has been recognized for some time, but it wasn’t until work in Europe (by the Austrian Met Service and by Manola Brunet here at URV) which reconstructed the older locations and took modern parallel measurements that adjustments can be made. Depending on location, temperatures before about 1880 can be up to 0.5 to 1.0C too warm because of this exposure effect.

A third possible bias is urbanization, due to the growth of cities around observing locations. I will return to this towards in a couple of slides with an example from London.

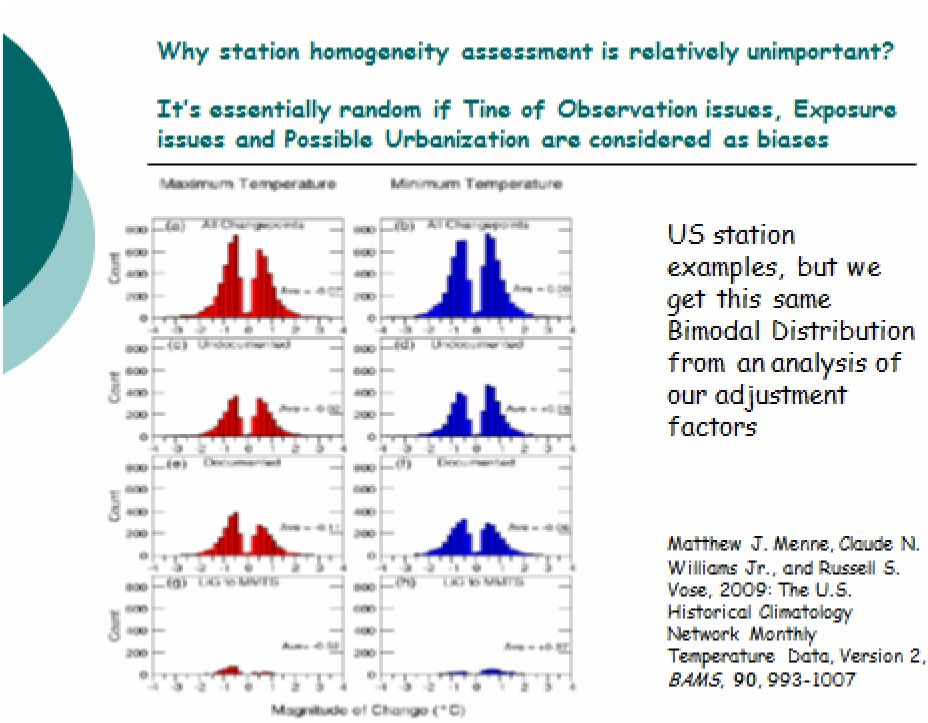
The two main ‘biases’ must be adjusted for first, before looking at the homogeneity of individual temperature series.



Here are some photographs of early Stevenson-type screens – in a comparison exercise at two locations in Spain and another at a site in Austria. The effects of these problems are mainly evident across Europe, but the changes (around the 1880s) occurred too early for many other parts of the world. This problem is important for some Australian states prior to 1910.

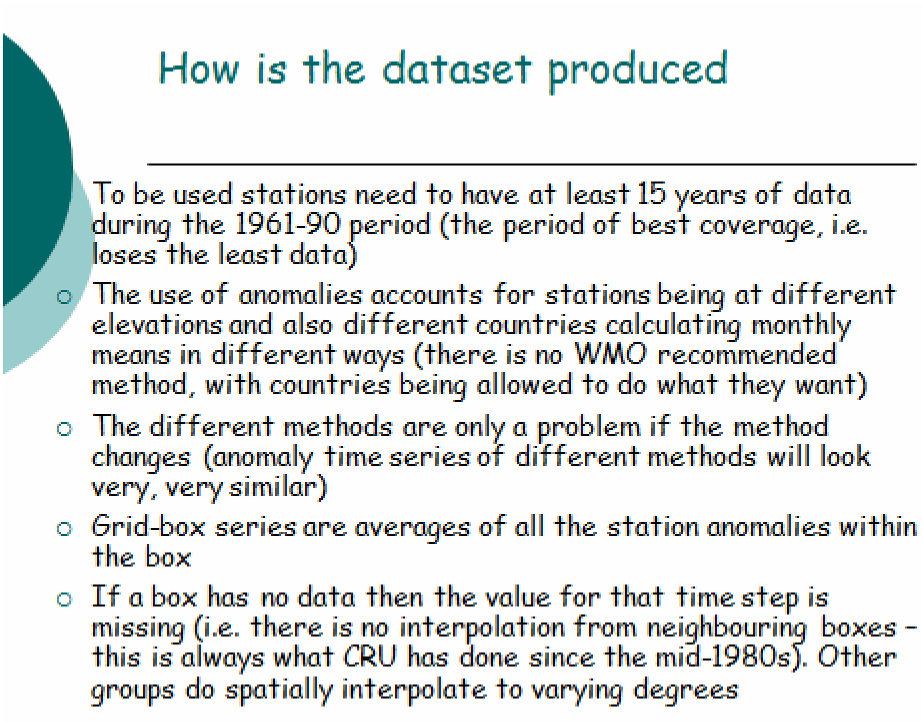


The other important bias is the effect of urbanization on temperature records. Some people think that temperatures measured in the centres of cities are strongly influenced by the growth of the city. This can be easily shown to be true, but NMSs don't locate sites in city centres – instead they were fully aware of these problems and sited instruments in parks and more recently at airports. Here I will show the example of London. There is only one long site in central London in St James's Park – this is quite close to where the beach volleyball was held during the Olympics. The site is just inside the park, near its NW edge – so quite lose to Horse Guard's parade. The different panels show the maximum, mean and minimum temperatures at the site – the orange line. These values should be compared with the two green lines, which are rural sites SW and NW of London (Wisley, a botanical gardens and Rothamsted, an agricultural research station, respectively). Over the course of the whole period from 1900, the St James's Park site has warmed at exactly the same rate as the two rural sites. This central London site is about 1 deg C warmer than it should be, but this extra warmth was attained prior to 1900 and likely over many centuries before. The St James's Park data could be used (we don't though) as in terms of anomalies from 1961-90 the anomalies barely differ from the two rural locations.



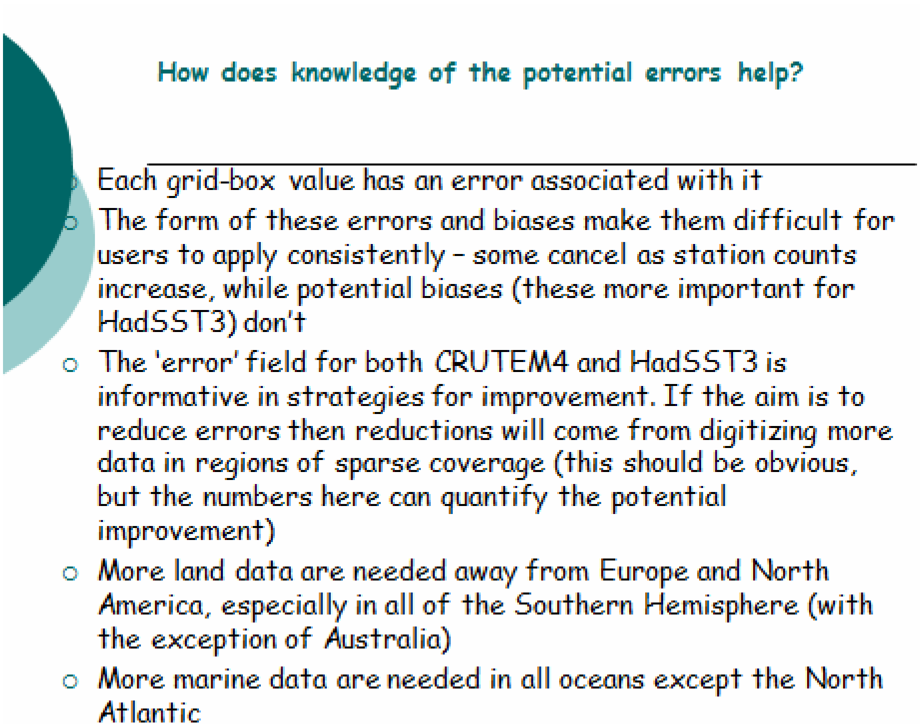
I mentioned earlier that homogeneity adjustments (ignoring the TOB and exposure issues) tended to be random. Here I show histograms of adjustment amounts produced by NCDC scientists for stations across the USA. The histograms show the number of adjustments of a certain size. They show typical bimodal distributions. These come about as adjustments near zero are difficult or impossible to find, so there is a 'missing middle' to the Gaussian shape that gives the bimodal distribution. The reason that the adjustments have little effect at the large scale is that the overall average of all adjustments is nearly zero. We produced similar plots for our 1980s adjustments in 2006 and got very similar bimodal distributions.

So while the effect overall is zero, they are important to do as researchers don't just want the large-scale averages, but also want to use the individual grid-box series as well. So every series has to be assessed for homogeneity. The aim of developing of a gridded product is not just the development of the global average.



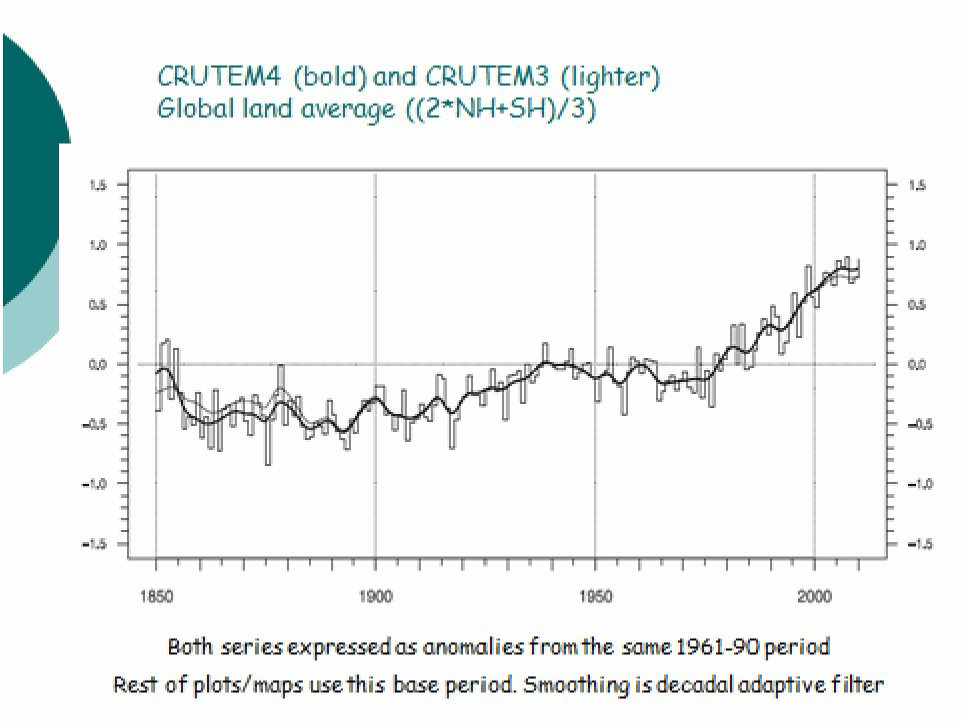
Everyone knows that temperature decreases with elevation, so in developing a gridded dataset it is not possible to use absolute temperatures. Grid-box series will be affected by the presence or absence of high- or low-elevation stations. The simplest way to overcome this is to remove the average temperature creating what are referred to as anomalies or departures from a common period, which is generally chosen to be 1961-90.

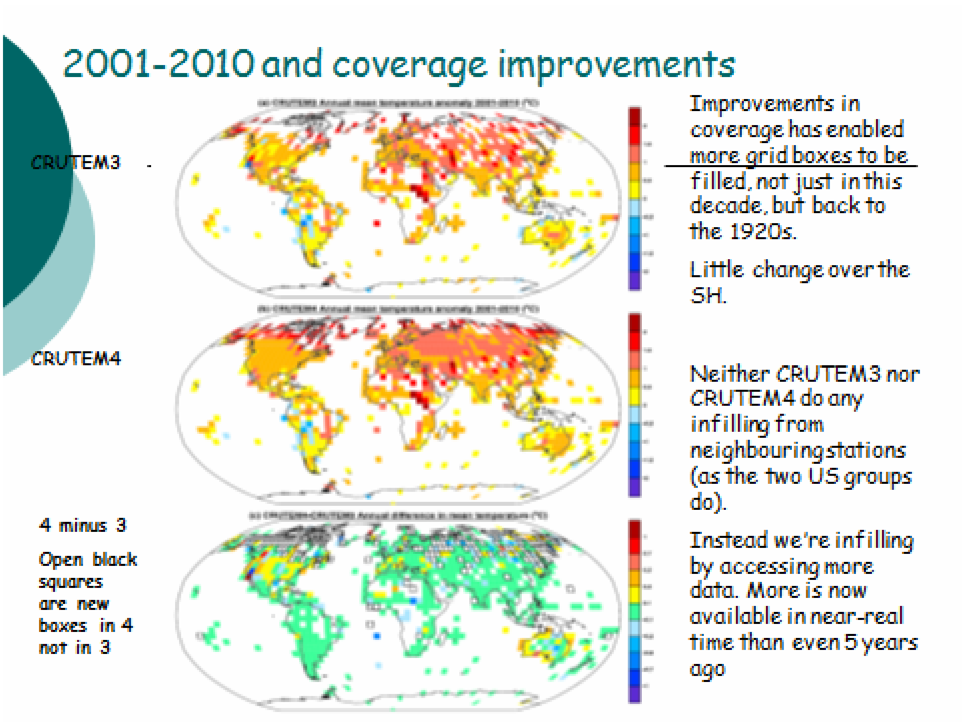
For CRUTEM4 we require that each station has to have at least 15 years for each of the 12 months of the year during the 1961-90 period. Stations that do not achieve this do not get used. Grid-box series are then produced by averaging the stations within each grid box – here we use one of 5 degrees of latitude and longitude. For our dataset we don't do any extrapolation of data – if a grid box has no station data then the grid box is set to missing. All other groups producing similar datasets (including the two in the USA) undertake some degree of spatial interpolation.



For every grid box dataset since 2006 we have produced an error estimate – for each box and also for large-scale averages. There are different components to the errors – the homogeneity of the station records, which because much is essentially random will cancel with more stations per box, the more pervasive biases (such as TOB, exposure and urbanization) which will not cancel and the sparseness of coverage in earlier years (which will lead to more missing values) which again increases the errors.

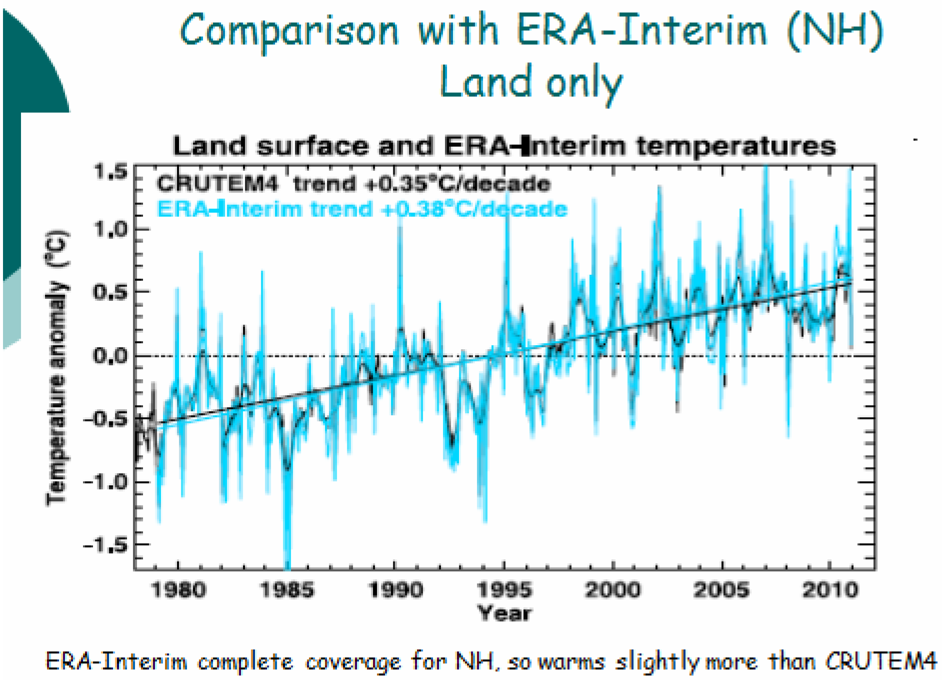
I won't go into detail about the errors, but knowledge of its structure is very useful for determining strategies for reducing the error. The errors are reduced by getting more station series, but additional sites in much of the tropics and the Southern Hemisphere will be much more important than more sites in the USA and parts of Europe. The error is much more affected by the location of the stations and researchers who say they use large numbers of stations (say ten times what we do) don't fully understand the error structure. Using ten times more data over the USA hardly reduces the error at all. This is why I commented upon the recent Brazilian data release. Even though this is only from 1961 and only about 300 series, it is likely to lead to a reduction in SH errors once the data are introduced.





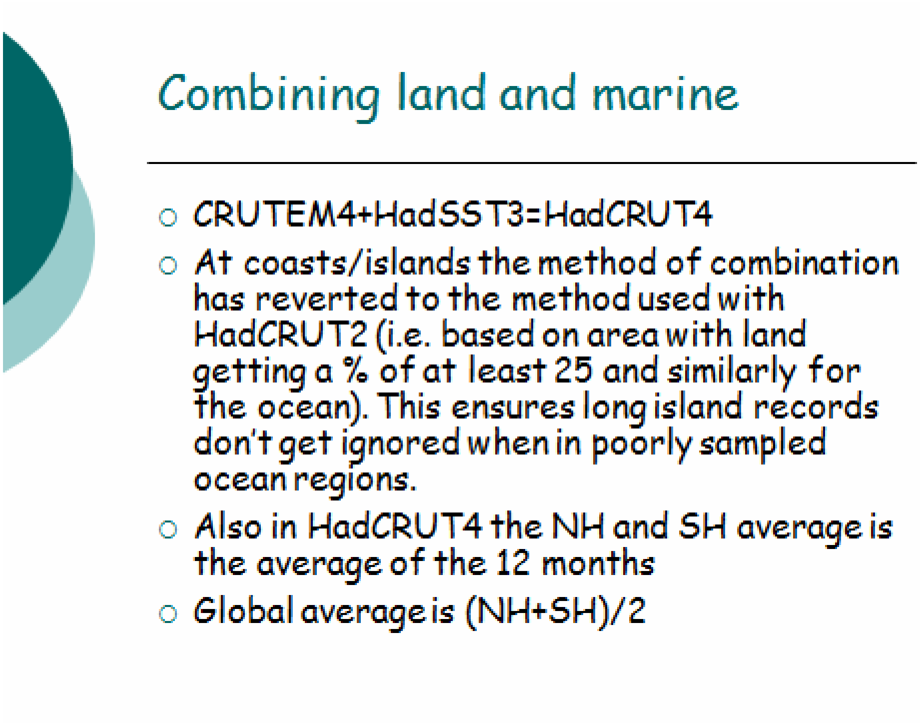
This slide shows the global average calculated for land areas, with the global average being the weighted average of the two hemispheres (with the NH counting twice and the SH once – roughly in proportion to reality). On this plot there is a comparison of the latest version for the land (CRUTEM4) with the older version (CRUTEM3 from 2006). Despite months of work adding in new data from NMSs and other sources and replacing all the US data with new US sources, there are only changes in the most recent decade (the 2000s) and before about 1880. This just illustrates how robust the land temperature record is! The additional data for Russia and Canada is the cause of the recent increase, as we now have more grid boxes with data and as these higher-latitude areas have been warmer recently, so the global average has increased very slightly by just less than 0.1 deg C.

The next slide shows the spatial coverage of the land data and the differences in coverage between CRUTEM4 and CRUTEM3. As we don't do any infilling, the only way to get more squares 'filled' with data is to find more data. Both Russia and Canada have made more homogenized series available in recent years. The differences between the two versions of the dataset over the US and Australia relate to the replacement of much of the data for these two countries with newer improved and homogenized datasets in the latest version CRUTEM4. The gaps in coverage in Brazil may be infilled, but there are larger 'white' areas in Africa and Indonesia where more work needs to be undertaken. Antarctica only has about 30 permanently manned stations – hence the few boxes across the continent. Automatic Weather Stations (AWSs) will improve this from the 1990s, but it will only lead to a doubling of the number of grid boxes. Satellite estimation here may be possible, but in some regions it is difficult to determine if the radiance temperatures that satellites measure are from the surface or from low-level clouds.

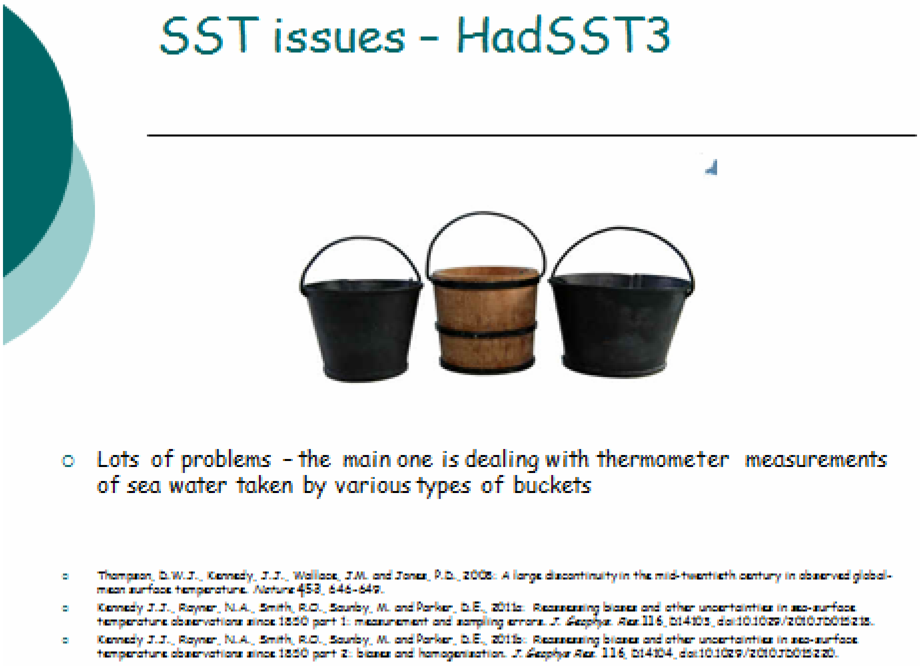


Later I will show comparisons of the dataset combined with marine data (HadCRUT4), but before doing this there is one dataset that is worth comparing with. Reanalyses are weather forecasts run with a consistent computer model and all the input data (satellites, weather balloons and surface data). The latest and best of these is called ERA-Interim developed by ECMWF at Reading in the UK. ERA-Interim runs from 1979 and in the plot I am showing the ERA-Interim average for the NH land regions compared to CRUTEM4. As you see the agreement is excellent at the monthly timescale. Some of the warm and cooler periods relate to El Nino and La Nina events during the last 30 years. Also the cooling from the Mt Pinatubo eruption in June 1991 is clearly evident. The two datasets agree on the trend. ERA-Interim warms slightly more, possibly as for this dataset the land areas are complete, so have no missing land-box series in the Arctic regions.

We have investigated using equal-area grids, instead of the latitude/longitude boxes, but the resulting gridded product will have less use to climatologists used to working with latitude/longitude grids. Most climate model and regional climate model output is available on latitude/longitude grids or is transformed to this before being released for use.



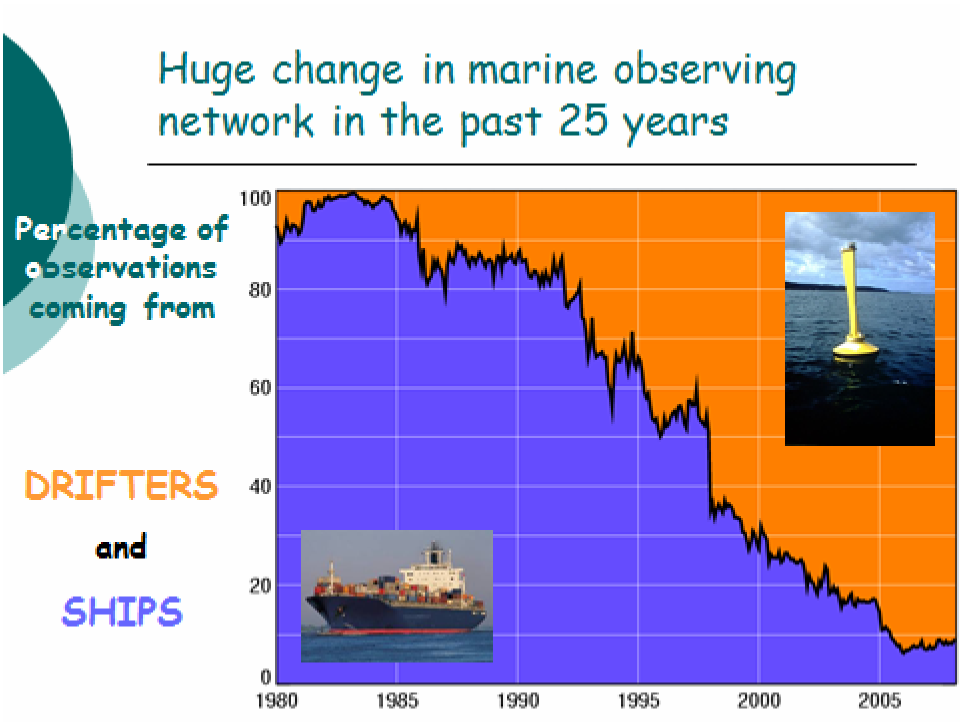
This slide introduces the dataset HadCRUT4 which combines the land data (CRUTEM4) with marine data (HadSST3). HadSST3 is a dataset of sea-surface temperatures across the world's oceans. When combining the two, a decision has to be made for grid boxes that contain data from both land and marine regions. In the latest version we average the two with weights of the land/ocean component, but ensure that the weight of land/ocean must be at least 25% and the other cannot exceed 75%. The purpose of this is to ensure that isolated islands in data sparse marine regions are not ignored in the combination. We have experimented with various methods of combination and HadCRUT4 is very robust to the particular method chosen.



HadSST3 is a dataset of sea surface temperature (SST) data, complimentary to the land dataset (CRUTEM4), and developed by my colleagues at the UK Met Office Hadley Centre. As you might expect there are also numerous problems with SST data and also with marine air temperature (MAT) data taken by ships. First, MAT data are much more variable within a month than SST, as SST doesn't vary much within a month nor little between day and night. If MAT data were to be used we would need 2-3 times as much data to derive averages as accurate as SST, so SST anomalies are used as surrogates for MAT for the ocean areas.

SST data have homogeneity problems principally due to way the measurements have been taken since the 19th century. SST measurements were first taken with wooden buckets on sailing ships. These were replaced with canvas buckets when steamships were introduced. Later (around WW2) ships began to use temperature measurements taken using thermistors located in the engine intake pipes of ships – which are used to cool the engines. It turns out that the bucket measurements are between 0.3 and 0.7 deg C cooler than the engine intake sensors. A method of adjustment has been developed to correct the bucket values, which is based on the thermal properties of a canvas bucket. The most important parameter here is the air-sea temperature difference, but an average value for the location and season can be used. If adjustments for this were not undertaken the world would have warmed much more than it has and there would be a large difference between coastal/island SSTs and land air temperatures prior to about 1940 which would be completely non-physical. In anomaly terms, large-scale averages of land temperatures and SSTs should agree. All homogeneity work on both components (land and marine) is undertaken independently, so the two separate datasets are mutually supporting one another.

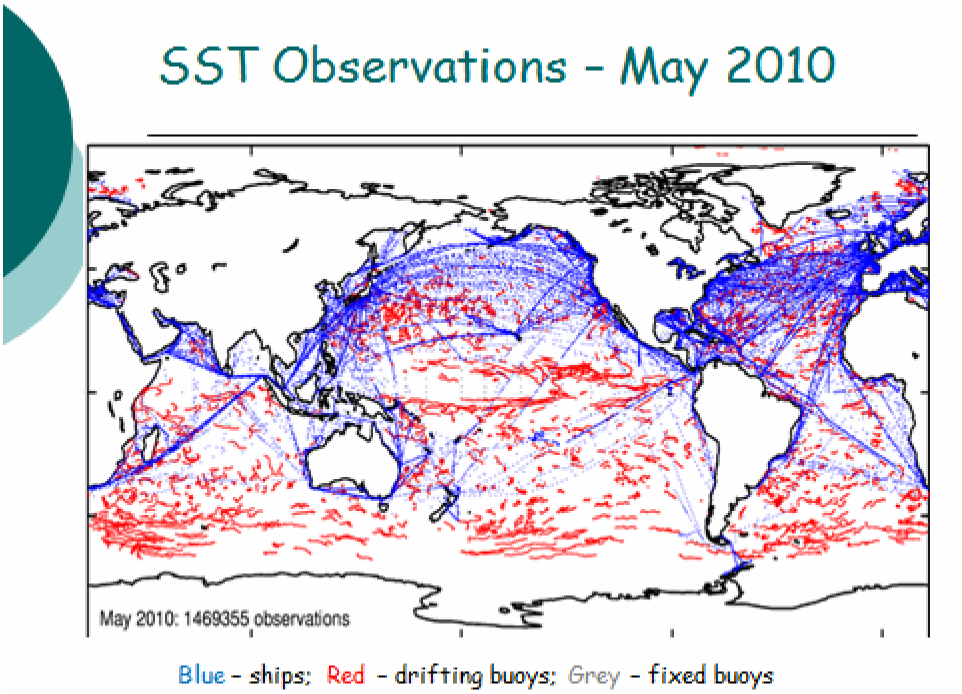
SSTs are vital for many other aspects of climate and weather services. For example, they are necessary for lower boundary conditions for all forms of weather forecasts and Reanalyses.

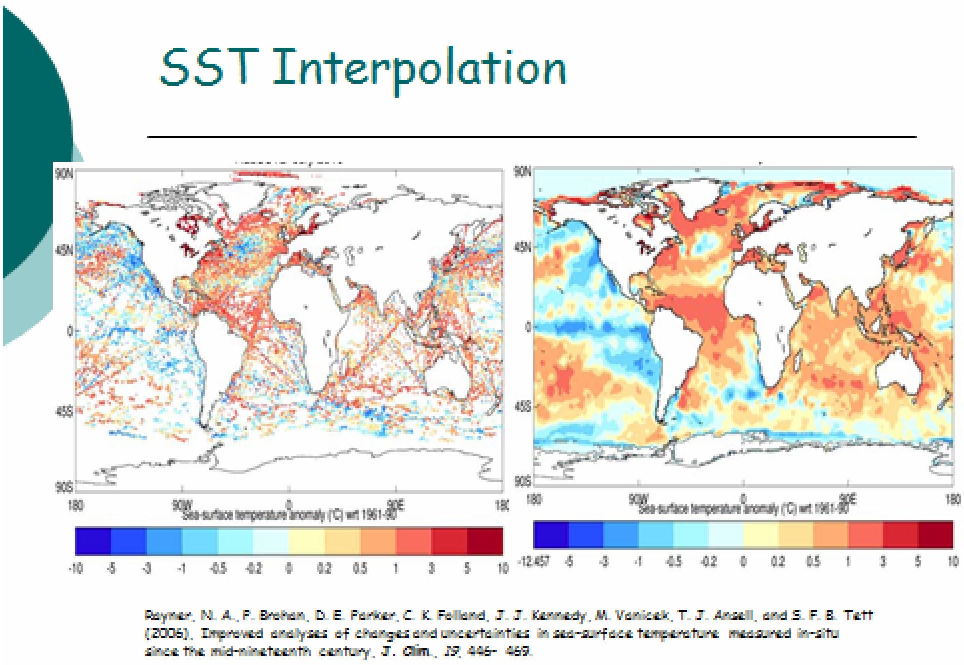


Despite the major change in SST measurements around 1940, there have been dramatic changes in recent decades also. Satellite estimates have proved very useful to expand coverage, but it has to be remembered that what the satellite sees is a skin temperature (top mm) while a ship measures SST several metres below sea level in the upper mixed layer. Also to improve weather forecasts, drifters have been deployed principally in the Southern Oceans and the tropics. These data are being extensively used for SST datasets, but they appear to measure SST about 0.1 to 0.2 deg C cooler than the ships. It is probable that the ship-based SSTs are slightly too warm because of the ship itself, so the drifters are the more accurate measure in an absolute sense.

Finally, the number of ships taking SST measurements is reducing because some shipping companies no longer want to take these measurements and there are security concerns. Much fishing fleet data (principally from South Korea, Japan and Taiwan) enters the dataset some years in arrears as the fleets don't want their positions to be known in real time.

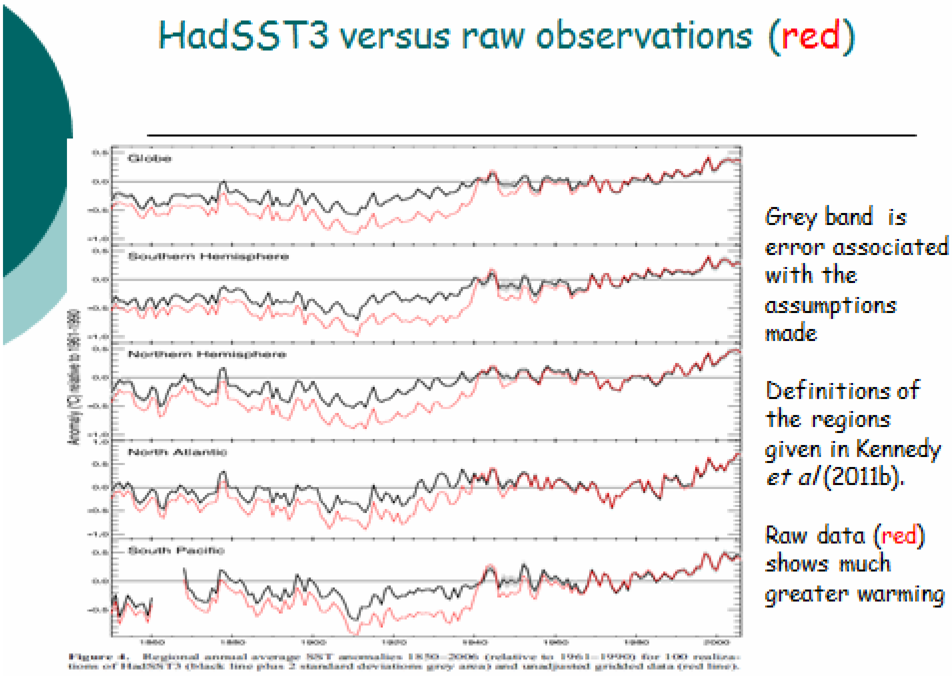
This slide shows the dramatic shift in marine recording over the last 25 years. In the 1980s most SST measurements came from ships, but now most come from the drifters. The drifters need to be continually deployed as they only have a lifetime of about a year. Also as the 1961-90 base period is based on ships, the recent SST anomalies could be biased low if the drifters are measuring 0.1 to 0.2 deg C cooler than the ships.



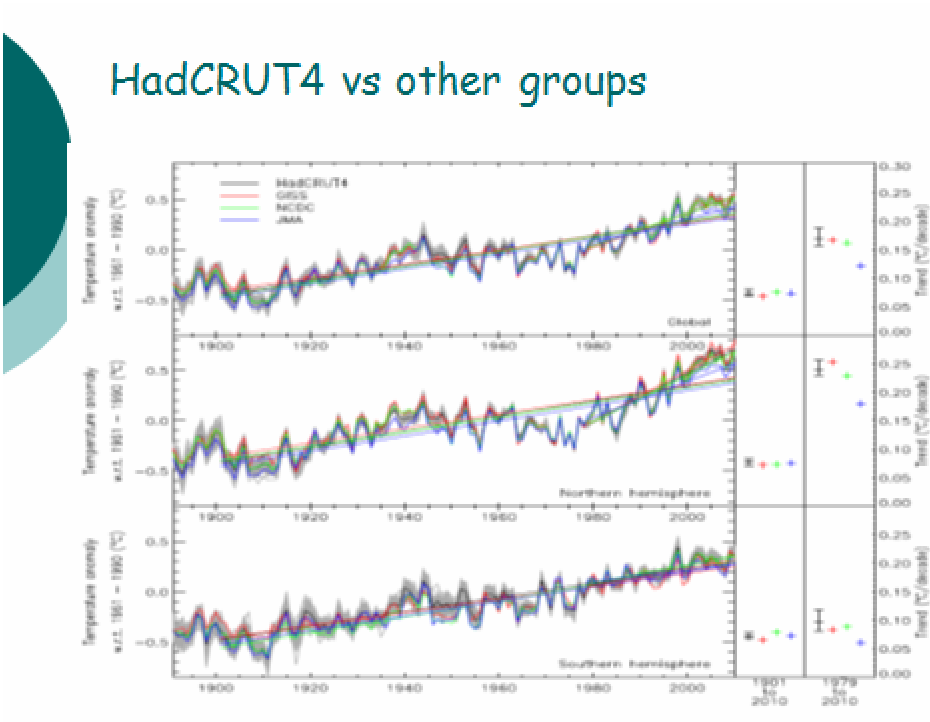


These two slides show how the HadSST3 dataset is put together for the month of May 2010. The first slide shows measurements coming from ships, drifters/buoys and also from fixed buoys in the tropical Pacific (set up to help forecasts of El Nino events).

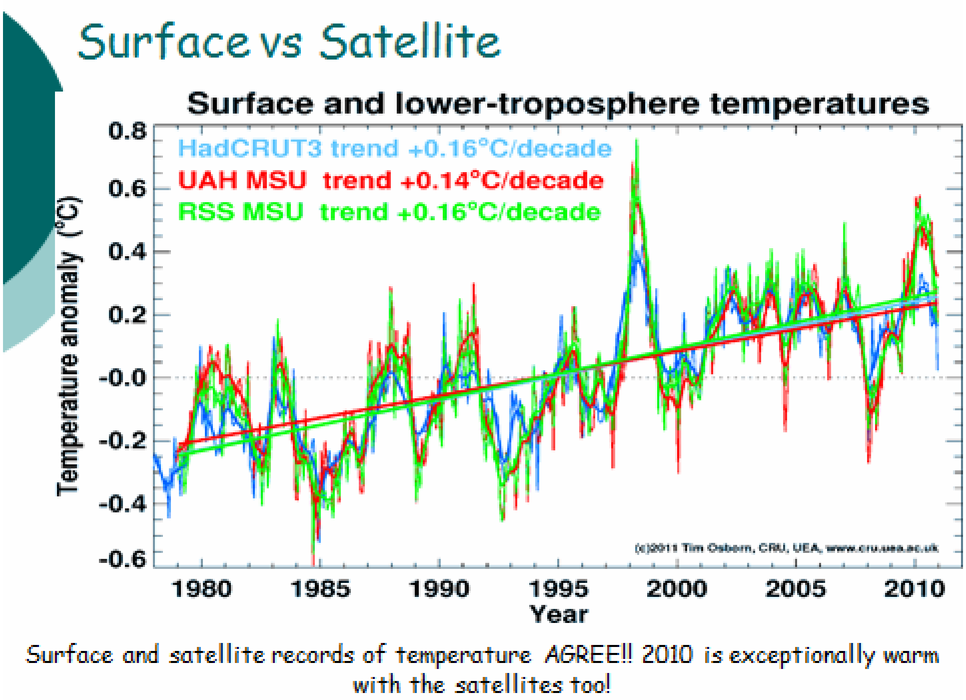
The second slide shows how the raw measurements are converted to anomalies (from 1961-90) then combined to produce anomaly maps for most areas of the world's oceans. There are still gaps though mainly in the Polar oceans, especially near areas of sea ice. As sea ice disappears in Arctic summers, SST data are appearing from ships travelling there, but these are difficult to use as there aren't data for the base period of 1961-90.



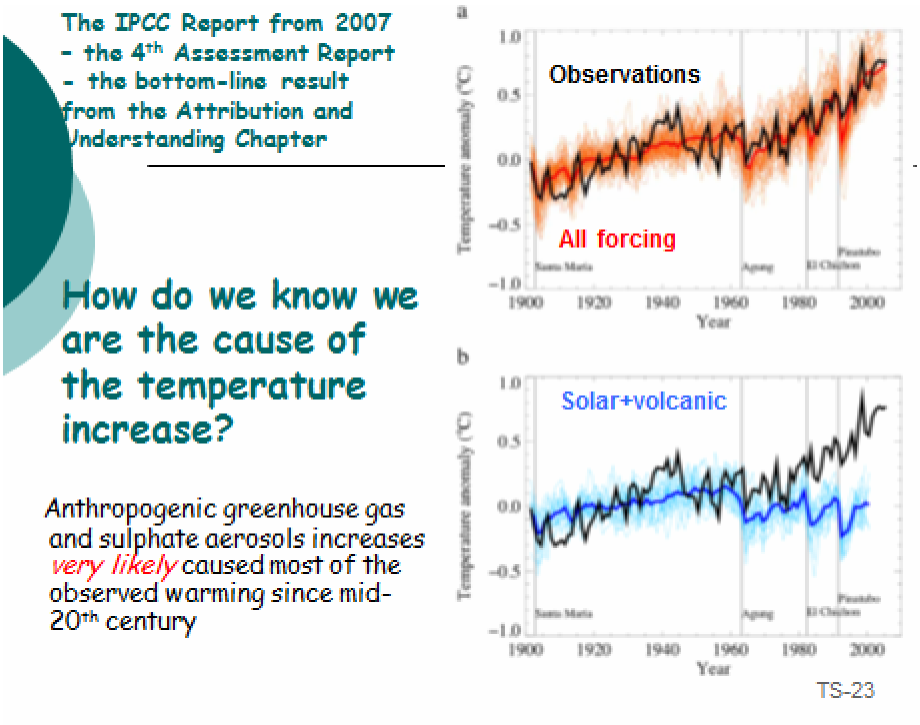
This slide shows SST averages from HadSST3 for regions of the world's oceans. The black lines are after adjustments for bucket/intake issues. The red are the raw values, if no adjustments for the homogeneity problems (principally the buckets before 1940) are made. If these adjustments were not made the warming of the world would be much greater. The adjustments to SSTs are by far the largest adjustment to any component of the global temperature averages. The sign of this adjustment is negative. Most people who work with surface temperature data only consider the land, but the SST data have larger adjustments and they affect a much greater percentage of the Earth's surface. The grey areas in the 1940s and 1950s represent revised SST averages after determination that British ships continued to use bucket measurements after WW2 up to about 1960 –because that is always what they had done, and no-one told them to use the engine intakes!



This slide shows comparisons of HadCRUT4 with two similar datasets developed in the USA by GISS/NASA and NCDC/NOAA and one from Japan (Kobe). Although all data groups make similar assumptions and use similar datasets, the agreement between them is excellent. The two right-hand panels show the trends (with error ranges for HadCRUT4) for the periods 1901-2010 and 1951-2010. Warming has increased over the last 80 years compared to the last 110 years. For the SH the linear trend is a good approximation for the last 110.



This panel shows comparisons of HadCRUT3 with two estimates of lower tropospheric (centres at about 5km above the surface) temperature estimates from NOAA polar-orbiting satellites. These satellite data are totally independent from any surface measurements used in HadCRUT4, and the comparison is of surface temperatures with those 5km aloft. Again the agreement is staggeringly good. The satellite measurements show slightly greater variability on monthly and seasonal timescales than surface temperatures as would be expected.



Finally before my conclusions, I will discuss the most important aspect from the last IPCC Report in 2007. This is the issue of how we know the rise in temperatures is as a result of the build-up of greenhouse gases in the atmosphere. Here, we use many computer model simulations of the last 100 years from different modelling centres around the world. The thinner orange lines are all the models run with all known factors that we think affect the climate. The thicker red line is the average. In the top panel this is compared with the instrumental temperature record in black. In the bottom the panel (in blues) the same models are run with just natural forcings which is just the result of changes in the output of the sun and in the effects of major explosive tropical volcanic eruptions. The global average temperature is the same black in both. As a result of natural forcings the world should have cooled very slightly over the last 50 years. Including greenhouse gases simulates the warming that has occurred.

Conclusions

- Adding more station data in the Arctic warms recent years
- Understanding the error structure enables strategies for reducing uncertainties to be determined. Finding and digitizing more stations in unobserved parts of the world is the key
- Biases resulting from changes in observation screens and the time observations are made are much more important than station homogeneity
- Urbanization is a factor in many cities, but NMSs are aware of the problems and sites are generally located in parks or at airports
- Marine data also have numerous homogeneity problems, but they are assessed independently so are mutually supportive of the land data
- There is excellent agreement between the different groups undertaking similar analyses around the world – not just in the USA, but in Japan, China and Russia
- Totally-independent estimates of temperatures in the lower-part of the atmosphere from satellites are also in near-perfect agreement with the surface temperature series
- The only way to explain the warming is to include human influences on the composition of the atmosphere within our climate model simulations.

Thanks!

- URV
- Geography Department

- For the nomination for the Doctor Honoris Causa

Thank you all for the nomination for the award of the Doctor Honoris Causa