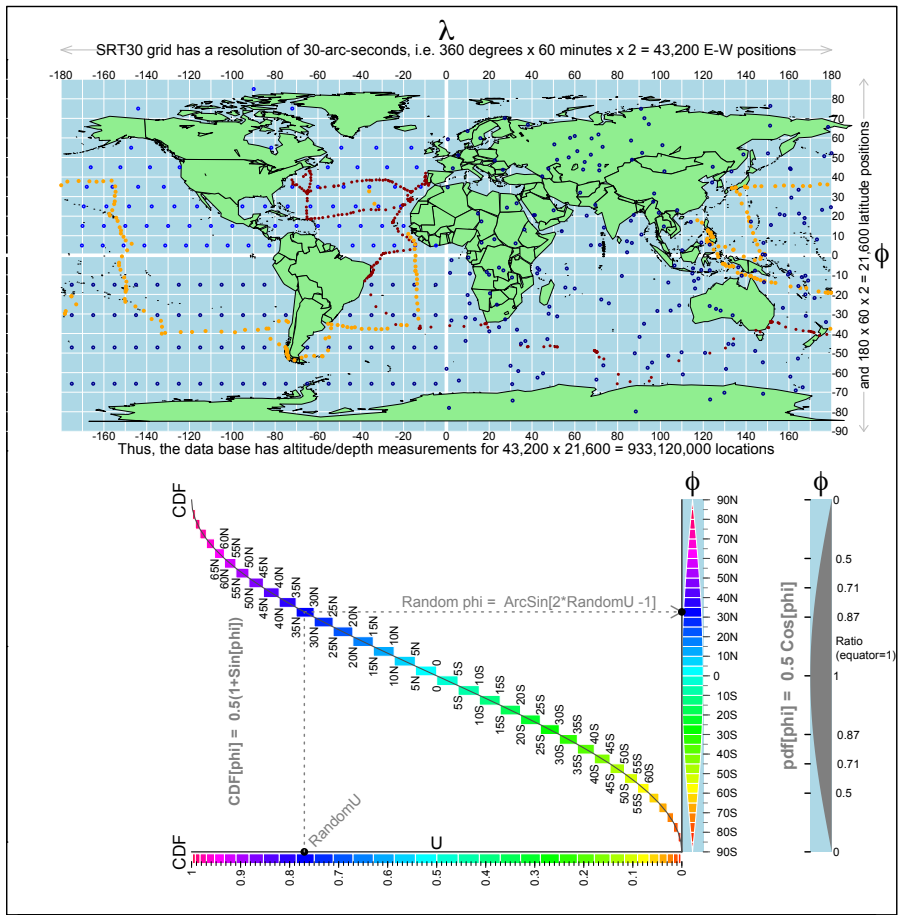

Figure 4: **A**. The resolution in the modern-day SRTM30PLUS database. Schematic representation of the rectangular grid of 933 million recordings in the SRTM30PLUS database, along with the locations of the soundings taken by the outward (red) and return (orange) portions of the 1872-76 Challenger Expedition. The soundings ranged from 4 to 4,475 fathoms: mean approx. 1400 (2700 metres, 1.6 miles). The locations, and the recoded depths, of all 500 soundings can be found online (see `http://19thcenturyscience.org/HMSC/README.htm`). The blue dots are for B.

B. (For the Data Mining Challenge) Some ways one might sample from the database to obtain a suitable sample of locations on the earth's surface. The sampling needs to reflect the fact that relative to the length of equator, the length of the corresponding 'line/circle' at latitude ϕ is Cosine[ϕ]. This function is shown in the 'segment-of-an-orange' shape displayed in the blue-background rectangles. In rejection sampling, one generates a ϕ value from U[$90S,\ 90N$], and retains it with probability Cosine[ϕ], i.e. as if a randomly selected location inside the rectangle shown at the bottom right 'landed' in the coloured area rather than the light blue background area. Another possibility is to sample ϕ directly, and without any rejection, from U[$-90S,\ 90N$], but to differentially weight observations by Cosine[ϕ]. Yet another is to use the 'inverse-CDF' method. The CDF function is best viewed by first rotating the Figure clockwise by 90 degrees; the inverse function is designed to be read in the 'as is' orientation, by (as shown with the dotted lines) entering the diagram on the horizontal (U) scale, and proceeding upwards and to the right to the vertical, (ϕ, latitude), scale. In effect, the method is equivalent to placing all the latitude lines 'end-to-end' and sampling uniformly from this concatenated 'line.' The sequence of small rectangles in the Figure is a necessarily-coarse version of this, whereas the smooth inverse of the smooth CDF curve (shows as a line) allows one to convert a random fractile value (i.e. $U \sim U[0,1]$) into a random latitude. The dark blue dots in A. in the grid representing the western hemisphere are doubly-systematic location samples – in the southern half, along equi-spaced longitude lines, and in the northern half, along equi-spaced latitude lines. The dark blue dots in the eastern hemisphere are locations whose longitudes were sampled from $U \sim U[-180, 180]$, and whose latitudes were sampled – independently of longitude – from the $[-180, 180]$ distribution shown as pdf(ϕ). [JH will remove the 'on land' locations].