SUPPLEMENTARY METHODS

Preparing storm track data and computing speeds. Best-track data are taken from IBTrACS31 (see "Data availability"). The approach followed by Kossin [1] was used with a few exceptions:

- The csv formatted data file was used instead of the NetCDF one.
- Duplicate reports for the same storm were eliminated. Eight storms had part of their tracks reported by the two centres, the US National Hurricane Center (NHC) and the Joint Typhoon Warning Center (JTWC). Based on the basin boundaries [1] the track that was out of bounds of a centres' jurisdiction was eliminated to avoid double-counting.
- There was no screening for non-main track types thus all types were included. As Kossin[1] points out, these can represent either physical or spurious cases. Kossin arbitrarily eliminated those with durations less than 3 days, which can potentially eliminate the spurious cases but also the physical cases. While non-main track cases account for only about 1.8% of all observations, only a fraction of these would be eliminated by the duration criterion.

The impact of these differences is minimal for the aggregate basins (GL, NH, SH). For the individual basins the character of the time series and trends are consistent, with trends I compute being typically slightly less in magnitude than Kossin's as expected by the use of a more resistant estimation method (see Supplementary Methods section "Miscellaneous statistical considerations") which is less sensitive to data at the ends of the record. In so much as the analyses here are focused on abrupt changes that persist from multiple years to a decade or more, the effects of such data differences will be inconsequential to the conclusions.

Bayesian information criterion (BIC) analyses. The methodology used to compute the BIC statistics follows that employed earlier [4] with a few points worth noting:

- Following established protocols I use the actual sample size (n) instead of the effective sample size (n_{eff}) in the BIC formula. The earlier use of n_{eff} [4] was an unjustified ad-hoc modification.
- I use only three models (TR, FS, SS) and do not consider the more complex, and less well-justified piecewise linear model.
- I do not consider autoregressive models (AR1 and AR2). Initial trials found that in all cases these produced more poorly performing models.
- I tested the model residuals and found that none were even remotely close to being significantly non-Gaussian.
- I use a different approach (see below) in assessing the relative validity of the three models considered (TR, FS, SS).

A more advanced methodology [5] than is typically used [4] was employed here to quantitatively compare the competing models, as opposed to the usual practice of simply reporting the "best" model as the one with the lowest BIC. In this approach weights are computed based on the Akaike Information Criterion (AIC) for each model. Here I adapt the methodology to use the BIC instead of the AIC. These two measures are closely related in that each consists of the weighted sum of two terms: one proportional to the mean squared error and the other proportional to the number of fitting parameters. The difference arises in that the BIC gives somewhat more relative weight to the latter "penalty factor" and thus tends to be more parsimonious in that it favours less complex models. In this way the analysis here is conservative in that it tends to err on the side of the trend model over more complex models.

The BIC weights that result from this approach are normalized to sum to 1 and are interpreted as relative likelihoods that each model is the best of the three competing models under consideration. For example, these likelihoods for the NH region (see Extended Data Table 2) are approximately 81% for FS, 14% for TR and 5% for SS. These can also be expressed as likelihood ratios. For example, a model based on change-points (FS or SS) is about 6 times more likely than one on trends (TR), as (0.808 + 0.050) / 0.142 ~ 6.

It is worth noting that the information-theoretic paradigm, from which the above are drawn, is distinct from the hypothesis testing paradigm. The information-theoretic perspective strongly frowns upon any interpretation of these statistics as relevant to the test of a hypothesis, which is considered inferior, and does not condone the use of the word "significant" in this regard [5]. There is no intent to assess the significance of any particular model or the differences between models. Instead, this approach is intended to be used to assess the relative merits of competing models each applied to the same data.

Generation of curves in Fig. 2. The values plotted in Fig.2 were based on aggregating data in latitude bins[14] and plotting the aggregates at the centre latitude of each bin. The bins used for both north and south latitudes are: 60-90^{o}, 55-60^{o}, 50-55^{o}, 45-50^{o}, 40-45^{o}, 35-40^{o}, 30-35^{o}, 25-30^{o}, 20-25^{o}, 15-20^{o}, 10-15^{o}, 5-10^{o}, and 0-5^{o}. The TCS climatology was computed by averaging all observations in each bin over 1949-2016. The changes in observations were computed by first creating a count of the number in each bin separately for the pre-satellite (1949-1965 or 1949-1981) and satellite (1966-2016 or 1982-2016) eras. The counts for each bin were converted to percentages based on the total number of observations in the given era. The values plotted in Fig. 2 as the change are the difference in percentages for a given bin, satellite minus pre-satellite era. Values for bins having fewer than 50 observations were not plotted.

Miscellaneous statistical considerations. For some calculations non-parametric alternatives to traditional techniques were used to guard against the effects of outliers. The "biweight mean" was used in lieu of the arithmetic mean, the "median of pairwise slopes" was used in lieu of least-squares regression, and the "Spearman correlation" was used in lieu of Pearson correlation [6].

These alternatives were employed in the BIC analyses, change-point determination, scenario change-point adjustment, reporting of trends, and assessing significance of trends. Although Kossin [1] repeated his trend analyses using the L1 norm as a means of mitigating the effects of outliers, by one common measure (i.e. the breakdown bound [6]) that approach has no resistance to outliers, as is the case for least-squares regression; in addition these methods are more sensitive to data at the beginning and ends of the data record. The mps approach used here has considerable resistance (i.e. a breakdown bound of 0.29) [6] and does not exhibit enhanced sensitivity to values at the ends of the record. Comparing my unadjusted trends (Extended Data Fig. 1, scenario 1) to Kossin's based on the L1 norm (Kossin's Extended Data Table 2), the results of significance assessments are the same, except for EP which I deem non-significant. However, the magnitudes of trends can differ.