Is LaTeX use correlated with the number of equations in a manuscript?

Althea V. Moorhead, NASA Meteoroid Environment Office
althea.moorhead@nasa.gov

Background and hypothesis

Astronomers often use LaTeX to compose and submit paper manuscripts (1), but its adoption is not universal. Some journals prefer Word submissions, and large organizations with standardized software loads may also pressure their scientists to use Word. Furthermore, text and table transcription has been shown to be more efficient with Word than with LaTeX (2).

LaTeX is most favored by fields that are mathematically intensive (1). While Knauff and Nejasmic (2) found Word to be similarly efficient in transcribing equations, users reported greater enjoyment of LaTeX. We suspect that Word may therefore influence its users to include fewer equations, but this underlying hypothesis is not easily tested.

We test the hypothesis that manuscripts submitted to the arXiV in LaTeX format have more equations on average than those submitted in PDF format only.

Manuscript selection and data collection

We decided to focus our investigation on manuscripts studying small bodies and those published in the journal *Icarus*.

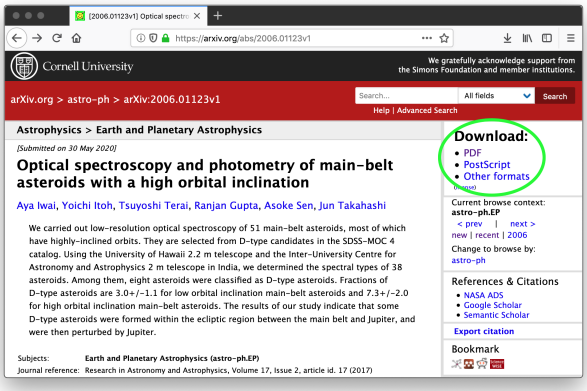
First, we use the arXiV API to select manuscripts that have a valid DOI and meet exactly one of the following criteria:

- the abstract contains "asteroid", or
- the abstract contains "comet", or
- the abstract contains "meteor", or
- the DOI corresponds to the journal *Icarus*.

Second, we use the NASA/ADS API to retrieve the journal name and whether it is considered peer-reviewed. We limit our analysis here to manuscripts for which published, peer-reviewed articles exist. We then keep only the first 500 manuscripts in each group.

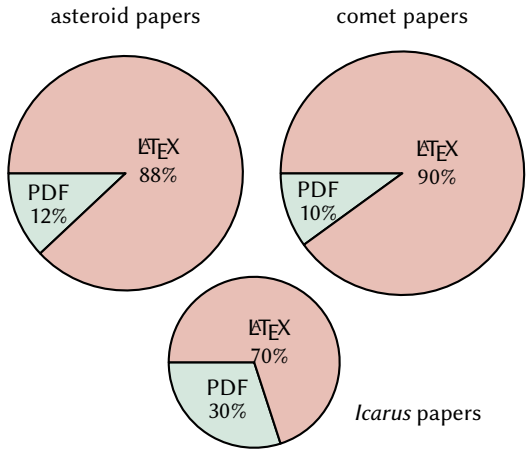
Third, we visually examine each manuscript and count the number of stand-alone equations in the main text, appendices, and supplementary material. If the authors choose to number their equations, we defer to their numbering.

We assume that the manuscript was composed using LaTeX if and only if "other formats" are available in addition to a PDF:



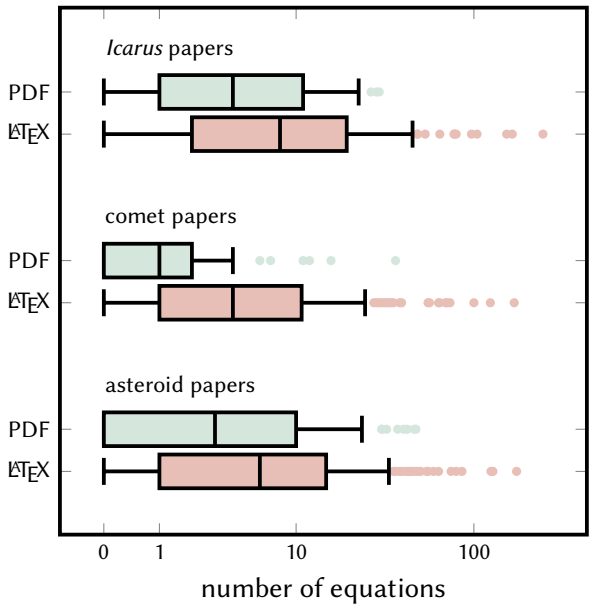
Data summary

We have so far counted equations in 488 asteroid papers, 499 comet papers, and 216 *Icarus* papers. We have not yet analyzed the small ($\sim$ 100) sample of meteor manuscripts available.



Preliminary results

We use a classic Tukey box plot to compare the PDF-only and LaTeX papers in our three samples. The box spans the 25th to 75th percentiles, with the median marked by a vertical black line. Whiskers mark the lower and upper adjacent values, and data points that lie outside this range are shown as points:



PDF-only manuscripts on the arXiV have roughly half as many equations as those with LaTeX source files.

References

(1) Pepe, A. et al., 2017, arXiV:1709.07020
(2) Knauff, M. and Nejasmic, J., 2014, *PLoS ONE*, 9, 1-14