



What's
The
Story?

Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number
University of Vermont, Fall 2023
Solutions to Assignment 20

“Permission to yell in a bloodcurdling way, sir?”

Name: Kevin Motia

Conspirators: Chris O’Neil

1. (3, 3, 3)

Using the main text you chose at the start of the semester, plot happiness time series in the following ways using the labMT lexicon.

The labMT word list was published with Ref. [?] in 2011, and has been occasionally upgraded to accommodate major changes in language use.

See and for the current version.

- (a) Process (destroy) your text so that it is a simple text file with one 1-gram per line—a vector of 1-grams.

To the extent possible, keep punctuation in as separate 1-grams. Periods, commas, semicolons, em dashes, ellipses, . . .

You can submit this as a separate file, but okay to just say you’ve succeeded.

- (b) First use the full lexical lens provided by labMT.

Make a single figure containing a stacked set of 7 plots with text windows of size $T = \lceil 10^z \rceil$ for $z = 1, 1.5, 2, 2.5, 3, 3.5$, and 4.0.

Stacked here means separated and stacked vertically, as opposed to directly overlaid. See examples for Moby Dick at the end of this assignment.

The notation $\lceil \cdot \rceil$ means round to the nearest integer.

- (c) Choose a ‘good’ text window from above, and repeat the analysis with lenses which exclude the central words around the neutral point.

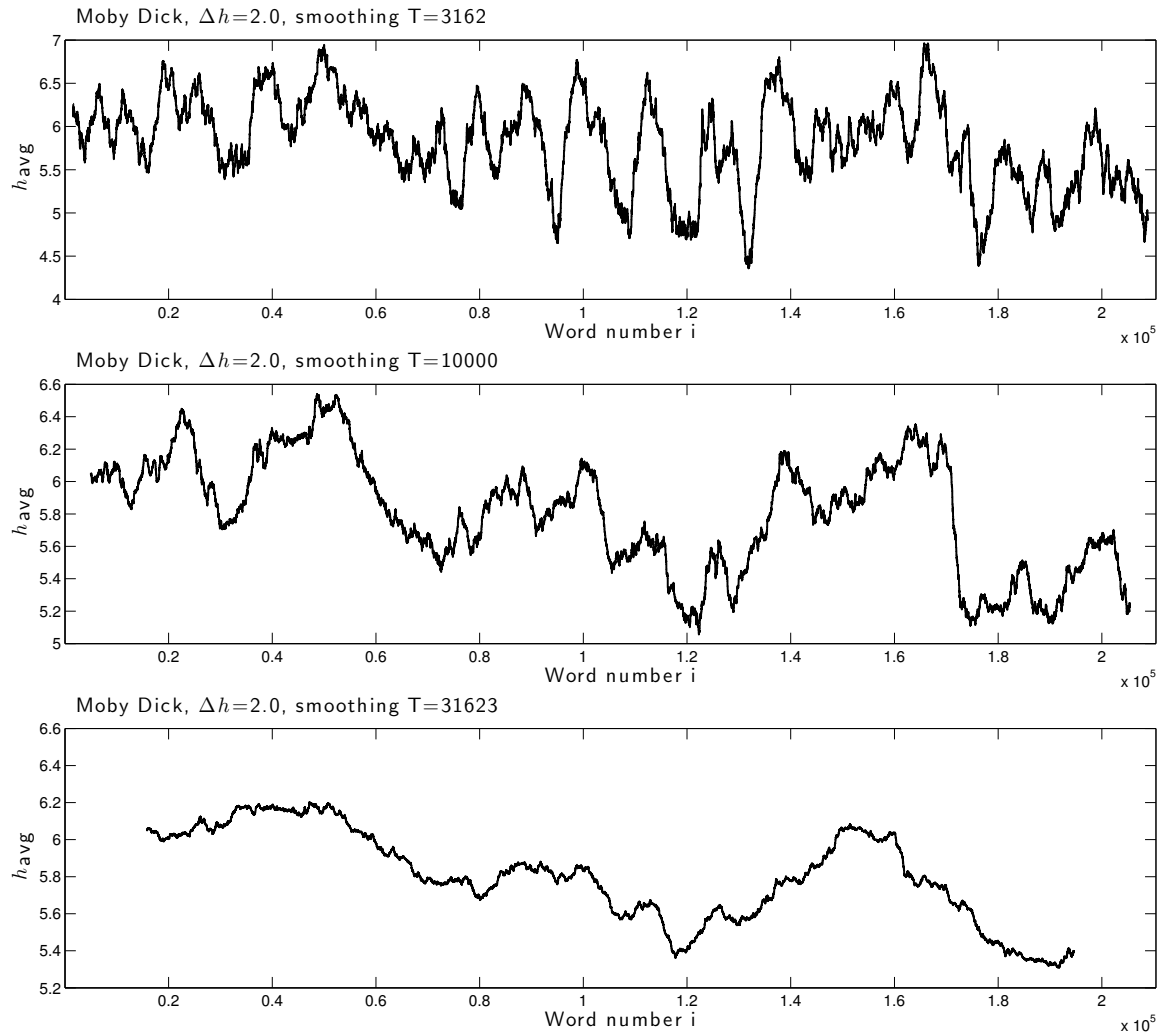
The blocked words are $h_{\text{avg}} \pm \delta h_{\text{avg}}$ where $\delta h_{\text{avg}} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$, and 3.5.

By ‘good’, we mean one that seems to you to produce a reasonable smoothing. Not too choppy, not too washed out.

Notes:

- The horizontal axis is “reading-experience time” corresponding to 1-grams in the text, running from 1 to N .
- The windows should overlap, sliding one word ahead each time. This is a simple averaging filter.
- Points should be located above the center of each window.
- So the point for the window running from n to $n + T - 1$ (T words) will be located at $n + (T - 1)/2$.
- Do not pre-filter the text for any given lens. Windows will contain variable numbers of words with and without happiness scores.

Three example averaging windows for Moby Dick with $\delta h_{\text{avg}} = 2.0$:



Solution:

Succeeded in destroying text so that it is a simple text file with one 1-gram per line – a vector of 1-grams.

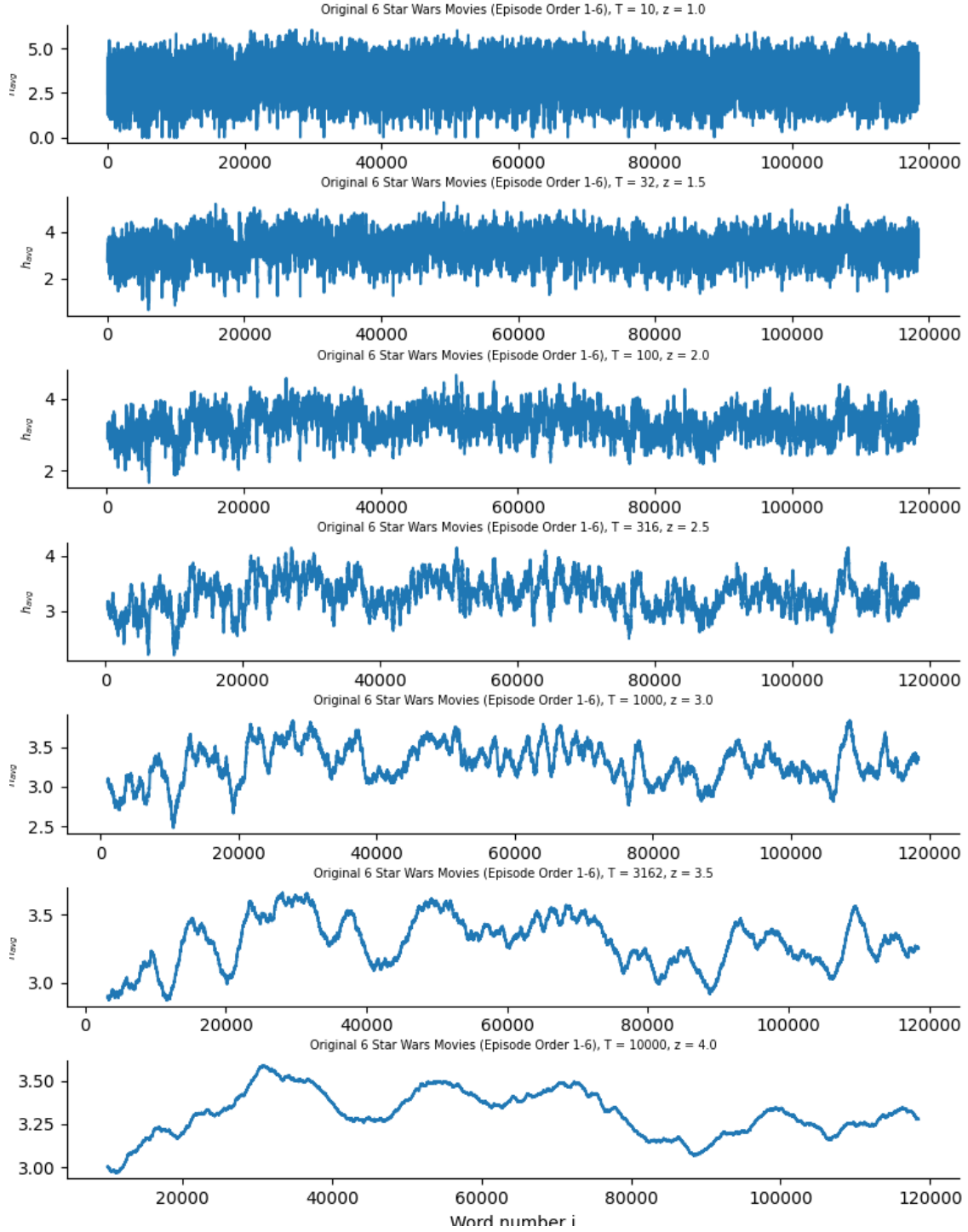


Figure 1: Varying window sizes of text size.

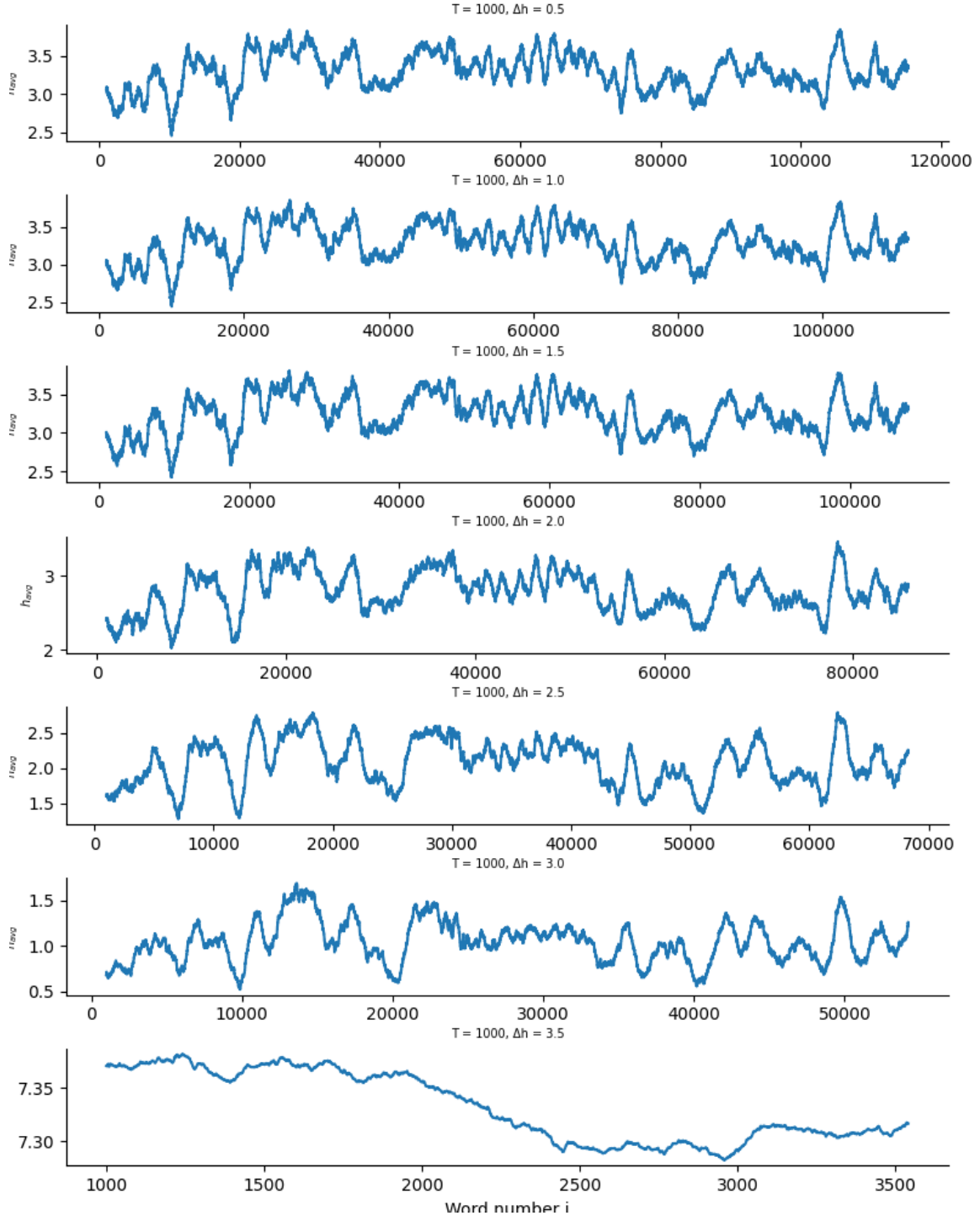


Figure 2: Constant window size with varying lenses that exclude central words around the neutral point.

