CSYS5040: Criticality in Dynamical Systems Assignment 3

The emphasis of this assignment is on using technical tools to explore the properties of a complex data set: collection, analysis, interpretation, and conclusions. You can include mathematical/theoretical analysis if you like, but it is not necessary and may not gain you any extra points. To achieve top marks you need to have chosen an appropriate data set, applied a variety of different tools to explore the data set, drawn appropriate conclusions, and (in the Summary section) explained the significance of your findings in a simple manner understandable to a work colleague. You may get higher marks for more readable and engaging content. **Your data needs to be a time series in order for it to be a dynamical system.**

Part 1:

Due date: Monday November 16th 2020, midnight, beginning of week 12,
Submit via: Video submission in Turnitin
Weighting: 20% of final mark
Length: 5 minutes min, 10 minutes max (hard limits – deductions for over/under time)

You will present a brief (5-10min) summary of the data set you have chosen to analyse, why you think it may have "complex" characteristics, and a proposal for how you intend to approach the analysis of the data. If you have preliminary results of your analysis you can present them during this session.

Potential sources of data:

1. Using HCTSA's comp-engine https://www.comp-engine.org data repository https://www.comp-engine.org/#!browse to get data to explore.
2. "Synthetic" data: Simulate a time series using Mathematica algorithms I've shown you in class, there are many different ways to do this I haven't covered as well, this can include running chaotic dynamical simulations, discrete maps, and bifurcations.
3. Mathematica's library of data is vast:
   https://reference.wolfram.com/language/ref/FinancialData.html
   https://reference.wolfram.com/language/ref/CountryData.html
   https://reference.wolfram.com/language/ref/WikipediaData.html
   https://reference.wolfram.com/language/ref/WeatherData.html
   https://reference.wolfram.com/language/ref/EarthquakeData.html
   and see the references in these articles to see other data and tools.
4. You can source your own data, but you have to believe that it will have some **complex-like behaviour**, and you will need to **justify this belief in your representation**.

Part 2:

Due date: Monday November 30th 2020, midnight: the beginning of week 14.
Submit via: TurnItIn Assignment 3 section on our Canvas site
Weighting: 40% of final mark
Length: 2000 words max. (hard limit – you will be penalised for exceeding!)
Format: There is a Mathematica document in the style of a report posted on the Canvas Assignment 3 page you will use as a guideline to the style of the final report

The final report is due to be submitted into TurnItIn on the **30th of November**. It will be the implementation of what you proposed in Part 1 of this assignment. There is a Mathematica document

in the style of a report posted on the Canvas Assignment page you should use as a guideline to the style of the final report

Some examples: These are ideas I've considered but not tried myself, they may be easier or harder than I suggest:

1. **Simple:** Take the housing OECD data I analysed for the Germany conference and recreate the results I got, and then extend this work using some more tools from HCTSA and some further thought on how best to combine different measures of analysis. An intermediate extension would be to use the HCTSA toolbox (rather than a subset of tools) to extend this analysis. Then interpret the different results you get from the different measures of complexity in a coherent, sensible way that brings out the insights each tool provides.
2. **Simple:** Use the comp-engine website and its data repository to explore a data set. To get a good mark for this you will need to be able to show that you understand the data very well, precisely what the HCTSA analysis is telling you, and to be able to interpret your results in terms of the original data. Comparing multiple different real-world data sets as well as synthetic data sets is a good way to extend this to an intermediate problem.
3. **Intermediate:** Extract many different types of time series data from Mathematica, simulate the data using stochastic differential equation modelling we covered in class, then compare (measure) the complexity or non-linearity of your simulated data to your original data.
4. **Difficult**: You can take some text from Wikipedia and synthesis it into a time series using https://reference.wolfram.com/language/ref/SpeechSynthesize.html and/or https://reference.wolfram.com/language/ref/Spectrogram.html and then measure the complexity of the voice time series and compare it to the complexity of the typed word complexity using HCTSA tools. Do this for different articles covering topics of differing levels of complexity.

You can get a perfect score with simple problems as well as difficult problems, it's just like the Olympics in gymnastics: the degree of difficulty makes it technically easier to get higher scores, but it's more likely you will make a mistake somewhere, but if you choose something simple you will need to be nearly perfect to get very high scores.

Rubrics

Part 1

Pass:

1. Hasn't quite kept to time but within reasonable bounds,
2. Reasonable connection from one idea to the next in the flow of the talk,
3. Describes a potential data set, it's a time series, but no discussion of content,
4. Some attempt made to justify the data as being of interest,
5. Has only a moderate understanding of the methods they will implement,
6. The justification for the data being complex is present but unclear,
7. Can only answer elementary questions about the topic, data, purpose, and methodology.

Credit, as for a Pass but including

1. Time is within bounds,
2. Good use of slides/presentation (slides not cluttered, well structured, communicates ideas to supplement the talk),
3. Explains the background to the data, where it's come from, why it's interesting, who might be interested in this sort of data/analysis,
4. Has some justification for the complexity/criticality of the data chosen,
5. Has preliminary look at the data with some plots (1st law of data analysis: plot your data),
6. Can answer most questions in a competent fashion.

Distinction, as for Credit and Pass but including:

1. Is able to interpret the initial plots of the data clearly and in such a way that preliminary insights are demonstrated,
2. Has a clear analytical approach to how the next stage of technical analysis is to proceed, and makes clear the reasons for the approach adopted,
3. Has read the literature on this data set or a similar data set and uses this literature to justify their expectations regarding the complexity of the data,
4. Is comfortable answering questions on any aspect of the work.

High Distinction, as for Distinction, Credit and Pass but including:

1. Initial analysis of the plots includes insights that are then used to guide how the next phase (deeper, specific analysis) of the work is to be carried out,
2. A detailed, specific and rigorous plan for the next steps is presented along with sound justifications based on material covered in class or from the literature,
3. Has a sound grasp of the literature around the data set being used, cites the relevant literature, recent results, why these results are significant within both complexity/non-linearity theory and the applied field from the data came from (or if it's synthetic data the field(s) that would be interested in the analysis),
4. Preliminary results are presented, discussed, and a discussion of progress to date: what worked, what didn't, lessons learned.
5. Is able to answer any question and then extend the answer to cover broader topics and show your ability to integrate ideas.

Part 2

Pass:

1. Has kept approximately to the word limit, or it is a little too short (<1000 words),
2. Has an Introduction to the data, an analysis section, and a conclusion,
3. The English is readable,
4. The background material describes the data and where it came from, what field it's relevant to,
5. Some methods are applied to the time series data but with little comprehension of what they mean or how to interpret them,
6. The conclusion is a list of disconnected and unrelated results from the analysis with no clear insights apparent.

Credit, as for a Pass but including

1. Has kept below the maximum word limit and is above 1000 words
2. Well structured sections and subsections, appropriately titled, with a suitable flow from one idea to the next,
3. The English is suitable for a professional report,
4. Background to the data is presented early, familiarizing the reader with the data, with some preliminary plots,
5. The analysis of the data looks for criticality/non-linearity, uses some appropriate measures, reports the results in a sensible and readable fashion,
6. Interpretation of the results is clear, functional, and shows some understanding of the multifaceted aspects of the data.

Distinction, as for Credit and Pass but including:

1. The introduction is a clear, accurate summary of the data, the field(s) that are interested in the analysis, and why it's relevant for complex/non-linear systems theory.
2. Background to the data includes suitable plots of the data that communicate interesting aspects that suggest the non-linear nature of the data.
3. The data analysis sections demonstrate a clear understanding of the methods used and how to interpret the results from those methods. The tools used explore the multi-faceted aspects of the data in such a way that the non-linearities are made explicit to the reader. The presentation of the results in the form of tables, plots etc. are clear and appropriate
4. The summary and conclusions draw together the results in a coherent fashion.

High Distinction, as for Distinction, Credit and Pass but including:

1. Introduction includes a referenced review of the field from which the data comes (or if it's synthetic data the field(s) that would be interested in the analysis). This includes the relevance for non-linear/complex systems theory. The references are not to be included in the word count.
2. Background to the data includes suitable plots of the data that communicate interesting aspects that suggest the non-linear nature of the data. Comparative analysis with linear data or some other alternative that makes clear precisely why the reader should interpret these plots as non-linear or critical.
3. The summary and conclusions show further insight into the data set and suggest other areas for further analysis, justified by a discussion of the results of the analysis.