DSCI354-351m-451 Final Exam

Roger H. French, JiQi Liu

09 December, 2018

Contents

0.1 Final Exam (20 pts)

- Will be held Monday 12/17/2018 from Noon - 3 pm
- Comprehensive over the course
- This is open book, open resource test
- Done as Rmd file to turn in
- Zip up your Final Exam folder and turn it in on blackboard

Final exam questions
- OI Stats questions - Data Analysis: Tidying, EDA - Essay Question - Do a exploratory data analysis on Degradation of Transparent Conductive Oxides

You have a pdf of OIStats book

- In your 3-readings folder of your Repo

1 Hypothesis Test Diamond Prices (3pts)

OIStats 5.28 Diamonds, Part I.

Prices of diamonds are determined by what is known as the 4 Cs:

- cut,clarity, color, and carat weight.

The prices of diamonds go up

- as the carat weight increases,
- but the increase is not smooth.

For example, the difference between

- the size of a 0.99 carat diamond
  - and a 1 carat diamond is undetectable to the naked human eye,

	0.99 carats	1 carat
Mean	\$ 44.51	\$ 56.81
SD	\$ 13.32	\$ 16.13
n	23	23

Figure 1: diamond prices table

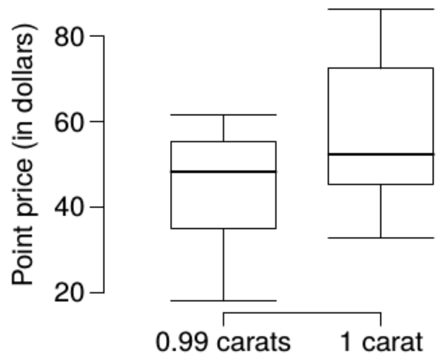

Figure 2: diamond prices

- but the price of a 1 carat diamond tends to be much higher
  - than the price of a 0.99 diamond.

In this question we use two random samples of diamonds,

- 0.99 carats and 1 carat,
  - each sample of size 23,
- and compare the average prices of the diamonds.

In order to be able to compare equivalent units,

- we first divide the price for each diamond
- by 100 times its weight in carats.

That is,

- for a 0.99 carat diamond,
  - we divide the price by 99. For a 1 carat diamond,
  - we divide the price by 100.

The distributions and some sample statistics

- are shown in Figure 1, and 2.

Conduct a hypothesis test

- to evaluate if there is a difference between
  - the average standardized prices
  - of 0.99 and 1 carat diamonds.

Make sure to

- state your hypotheses clearly,
- check relevant conditions, and
- interpret your results in context of the data.
- use good code style

	0.99 carats	1 carat
Mean	\$ 44.51	\$ 56.81
SD	\$ 13.32	\$ 16.13
n	23	23

Figure 3: diamond summary statistics

```
# Answer goes here:
```

2 95% Confidence Intervals for Diamond Prices (3pts)

OIStats 5.30 Diamonds, Part II.

In Exercise 5.28, we discussed diamond prices

- (standardized by weight)
- for diamonds with weights 0.99 carats and 1 carat.

See the table for summary statistics (Figure 3), and

- then construct a 95% confidence interval for the average difference
  - between the standardized prices of 0.99 and 1 carat diamonds.
- You may assume the conditions for inference are met.

```
# Answer goes here:
```

3 Visualizing Anscombe's Quartet of 'Identical' Simple Linear Regressions (3pts)

Four x-y datasets which have the same traditional statistical properties

- (mean, variance, correlation, regression line, etc.),
- yet are quite different.

The anscombe quartet is available in R,

- and they even give code to plot it.

But for this problem

- you should make your own code for visualization.

Using the Anscombe quartet,

- demonstrate the differences in these four datasets
- using different plots and approaches to visualize them.

Use ggplot for your visualizations.

```
# Answer goes here:
```

4 What is data science? (5 paragraph essay with citations) (4pts)

What is Data Science

- What do you find most interesting or exciting about data science and EDA?
- What defines data science and how has it come about.

- What are its characteristics, and what are the elements of
  - a data science tool chain,
  - a data science pipeline, and
  - a data analysis.

Use the structure of a 5 paragraph essay

- Introduction,
- 3 topic paragraphs,
- 1 concluding paragraph) with citations/references.

Essay Answer goes here

5 EDA of TCO degradation (7pts)

5.1 This problem will be similar to

This problem will be similar to

- Proj1 on Degradation of Hard Coat Acrylics.
- But you are given a csv file of a clean and tidy data set.

You will need to

- do EDA and
  - make figures
  - and summaries of what you find.

And list the insights you can develop from your EDA.

5.2 TCO's are transparent conductive oxides

TCO's are transparent conductive oxides

- Such as the materials ITO, AZO and FTO.
- They are used in displays, televisions, phones, photovoltaics etc.

Heather Lemire Mirletz did her MS thesis on TCO degradation

- And has a journal paper published.

Figure 4 is the abstract of the paper

Figure 5 is a mindmap showing the structure of her data science study

- there is a pdf of this mindmap in the tco-degr data folder.

Figure 6 gives information on the samples studied and the exposures used

Figure 7 is information about the exposure parameters used for the samples

Degradation of Transparent Conductive Oxides: Interfacial Engineering and Mechanistic Insights

Heather M. Mirletz[1,2,3,*], Kelly A. Peterson[1,3,5], Ina T. Martin[2,4,3], Roger H. French[1,3,4,**]

Case Western Reserve University, Cleveland OH 44106

Abstract

Transparent conductive oxides (TCOs) are a known failure mode in a variety of thin film photovoltaic (PV) devices, through mechanisms such as resistivity increase and delamination. Degradation science studies of these materials, as well as most PV systems, have primarily utilized industry standard qualification protocols, which are not designed to be used as lifetime prediction tests. This work applies a data science approach to this engineering challenge, utilizing commercially available TCOs and subjecting them to an array of stressors, including environmental and material stressors. Optical, electrical and surface sensitive TCO property metrics were monitored and analyzed en mass, broadly observing critical macroscopic material properties. Different degradation mechanisms and modes were observed when different stressor combinations were applied; TCO surfaces are sensitive to the proportion of water and light in an exposure, yellowing of the TCO only occurs when humidity and UV light are combined, and PEDOT:PSS (poly(3,4-ethylenedioxythiophene) poly(styrenesulfonate)) application results in hazing and roughening of AZO. Using multi-variate analytics and plotting crit-

*nee Lemire

**Corresponding author.

Email address: `roger.french@case.edu, Tel. (216) 368-3655` (Roger H. French)

[1]Solar Durability and Lifetime Extension (SDLE) Center

[2]Materials for Opto-electronics Research and Develeopment (MORE) Center

[3]Dept. of Materials Science

[4]Dept. of Physics

[5]Dept. of Chemistry

Preprint submitted to Solar Energy Materials & Solar Cells *February 13, 2015*

Figure 4: tco-degr abstract

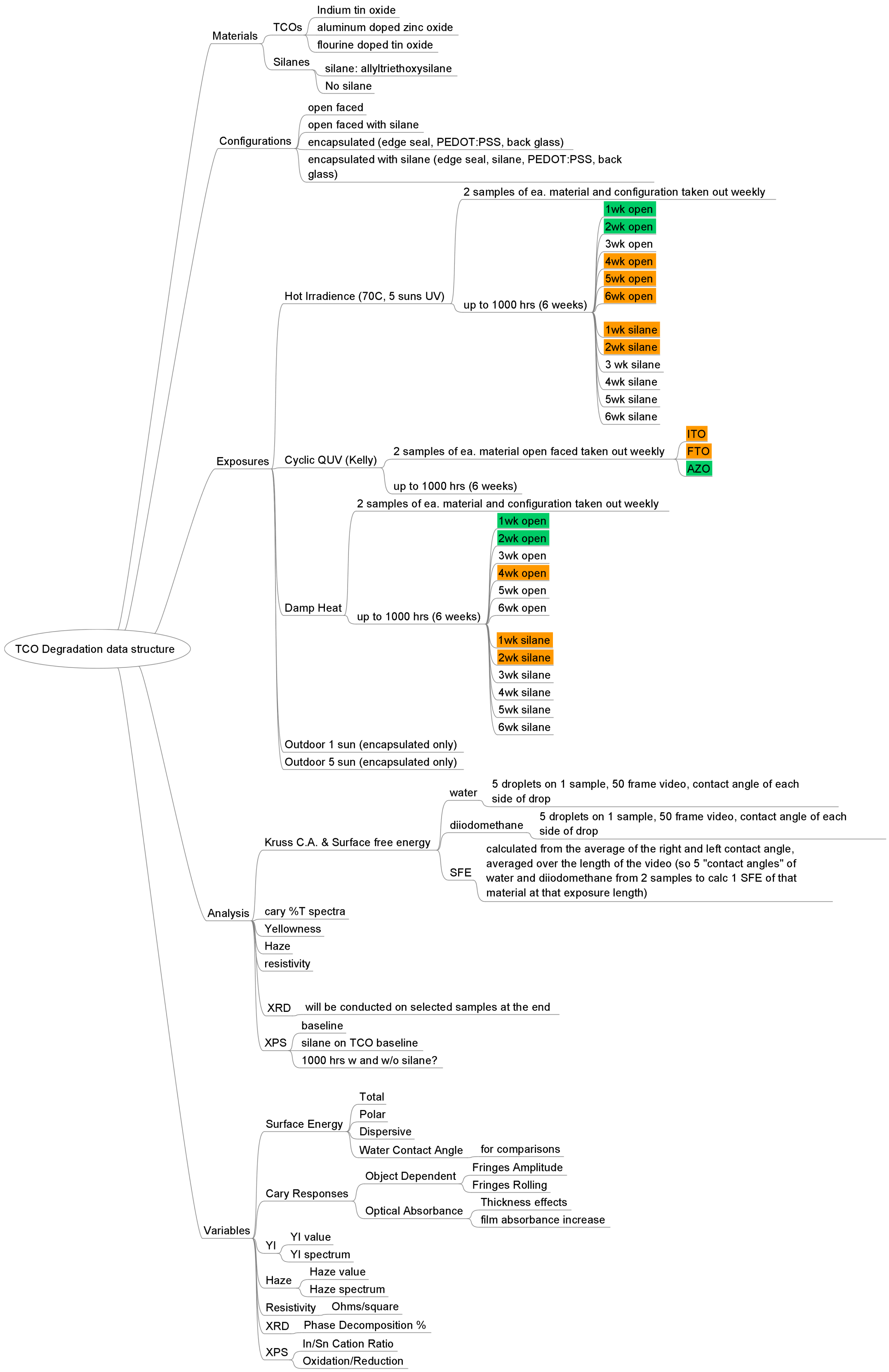

Figure 5: tco-DataStructue

Study Sample Map

		Bare TCO		PEDOT:PSS		Silane	
		AZO	ITO	AZO	ITO	AZO	ITO
Open Faced	Hot QUV	12	12				
	Damp Heat	12	12			12	12
	Cyclic	18	18				
Encapsulated	Damp Heat	6	6	12	12		
	1x Outdoor	6	6	12	12		
	5x Outdoor	6	6	12	12		

**encapsulated TCOs without PEDOT:PSS were removed from exposure in pairs at exposure steps 1, 3, and 6.*

Figure 1: Matrix showing the number of samples, exposure type and exposed sample configuration. Samples were removed, tested, and stored in a retained sample library at 6 time increments throughout the exposure duration.

Figure 6: tco-samples

2.2. Exposure Parameters

Five exposure protocols were used (3 accelerated and 2 real world): ASTM G154 X2.1-Cycle 4 ("cyclic"), Modified ASTM G154 X2.1-Cycle 4 ("5 suns UV"), IEC 61626 ("damp heat," "DH"), and outdoor exposures at 1x and 5x concentrations on a dual-axis tracker (see Table 1). Note that ASTM G154 X2.1-Cycle 4 is one of a commonly used set of irradiance settings; in this case, the irradiance that is approximately equivalent to 5 suns of UV light at 340 nm.[55] The 5x outdoor exposure used front surface aluminized hexagonal mirror concentrators to concentrate the sunlight, such that the final irradiance is 5 times the sun's direct irradiance.

Figure 7: tco-exposure parameters

Exposure Name	Conditions	Equipment	Stressors
Modified ASTM G154 X2.1-Cycle 4, "5 suns UV"	70 °C and 1.55 $W/m^2/nm$ @ 340 nm	QUV Spray	Heat, UV light
IEC 61646.10.13, "damp heat," "DH"	85 °C, 85% relative humidity	CSZ ZPH8	Heat, Humidity
ASTM G154 X2.1-Cycle 4, "cyclic"	8 h @ 70 °C, 1.55 $W/m^2/nm$ @ 340 nm, 4 h @ 50 °C with spray in dark	QUV Spray	Heat, UV light, Humidity
Outdoor 1x	Cleveland weather June-October	dual-axis tracker	"Heat", Sunlight, Humidity
Outdoor 5x	Cleveland weather June-October, front surface mirror concentrator	dual-axis tracker	"Heat", Sunlight, Humidity

Figure 8: exposure conditions

And a table about the exposure conditions

The tco-degr csv datafile is located

- in the tco-degr subfolder
  - of the FinalExam
- it consists of 771 observations of 23 variables

Some questions to try to address, showing your results.

- Which type of TCO (ITO, AZO, FTO) is most stable?
- Which type of Exposure is most aggressive?
- How do 'open' vs. 'encapsulated' samples compare.

What other insights

- can you identify and demonstrate
- from your EDA?

```
# Answer goes here:
```