E^3

EARTHQUAKE INTENSITY

USING DATA SCIENCE FOR BETTER UNDERSTANDING

ERIC HEPP

COMMUNITY DECIMAL INTENSITY USEFULNESS

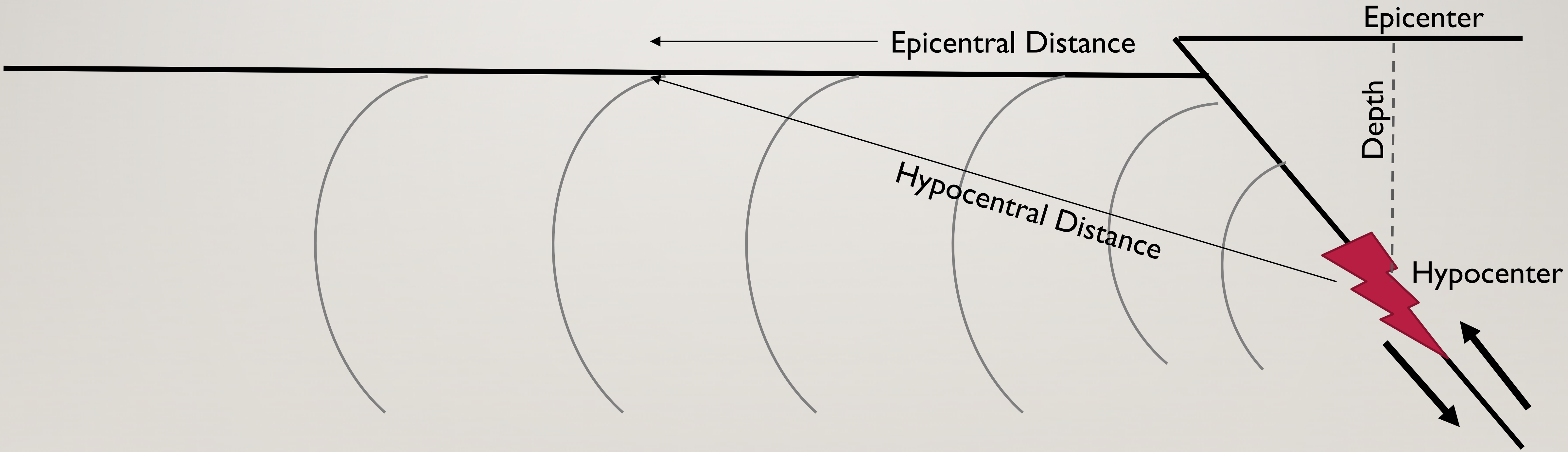
Conclusion: Based on this dataset, Community Decimal Intensity (CDI) is useful as initial evaluator; it is cheap and easy to collect data.

- It can be a reasonable predictor of Modified Mercalli Intensity (MMI) when considering the maximum values of each which occur near earthquake epicenter, especially when combined with additional factors of depth and region.

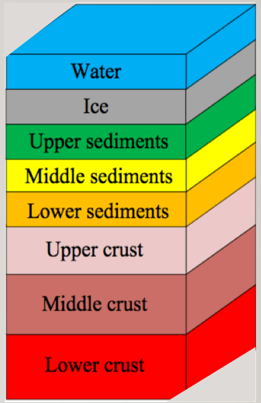
- CDI can be predicted from magnitude and distance when enhanced by additional non-traditional factors.

INTENSITY AND DISTANCE

SHAKING INTENSITY		I	II-III	IV	V	VI	VII	VIII	IX	X+
EXPECTED DAMAGE (Building Type)	(Sturdy)	none	none	none	V. Light	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy
	(Fragile)	none	none	none	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy	V. Heavy



INTENSITY FACTORS AND MEASUREMENT



Contributing Factors:

Magnitude | Distance | Depth | Crustal Properties

INTENSITY

GOAL: Predict it Better!

Methods to Quantify:

Modified Mercalli (MMI)

- Expert / Reviewed
- May employ surface motion sensors
- Modified and used in US from 1931

Community Decimal (CDI)

- Crowd sourced from DYFI reports
- Began 1991, in common use after 2001

INTENSITY MODELING: VALIDATE CDI

Magnitude

Distance

Depth

Crustal Properties

INTENSITY

Modified Mercalli (MMI)

- Expert / Reviewed
- May employ surface motion sensors
- Modified and used in US from 1931

Community Decimal (CDI)

- Crowd sourced from DYFI reports
- Began 1991, in common use after 2001

Model

INTENSITY MODELING: ADD DETAIL

Magnitude | Distance | Depth | Crustal Properties

Model

INTENSITY

Modified Mercalli (MMI)

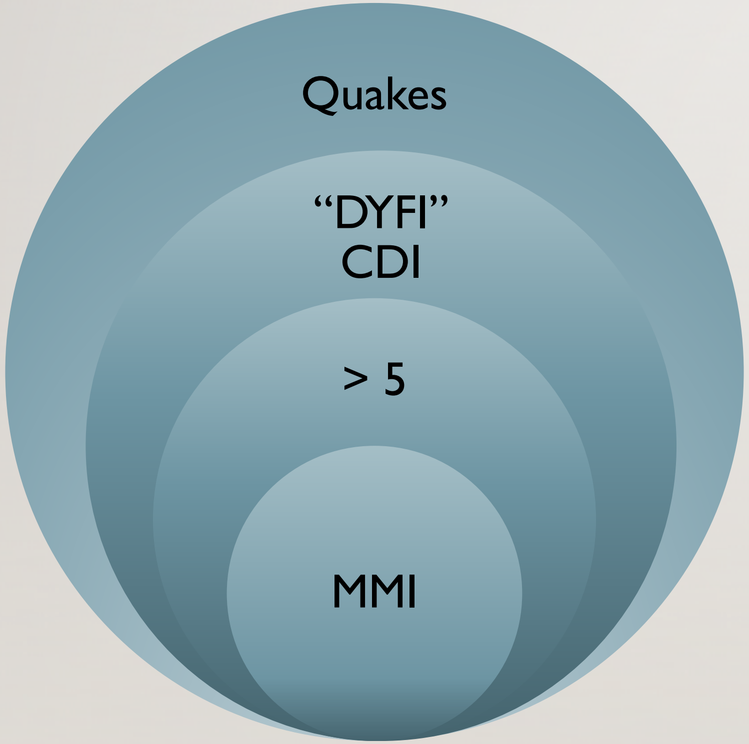
- Expert / Reviewed
- May employ surface motion sensors
- Modified and used in US from 1931

Community Decimal (CDI)

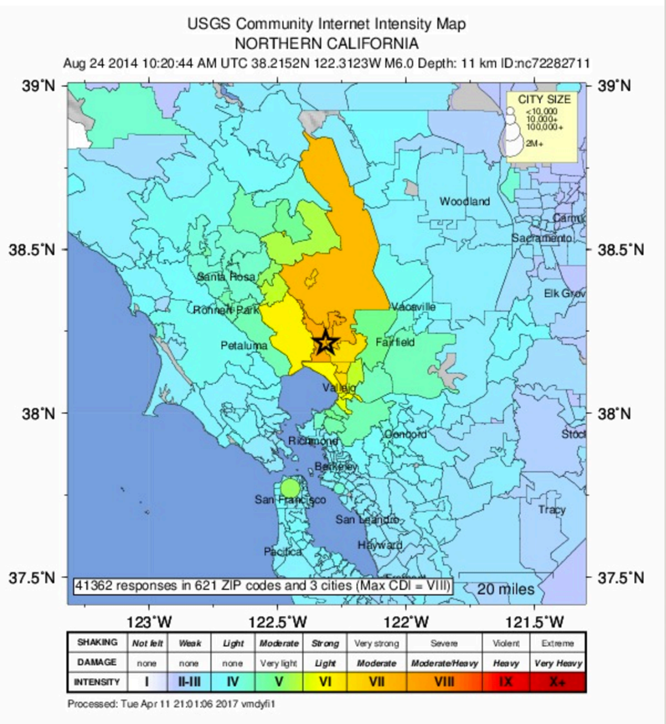
- Crowd sourced from DYFI reports
- Began 1991, in common use after 2001

DATA: HOW MUCH

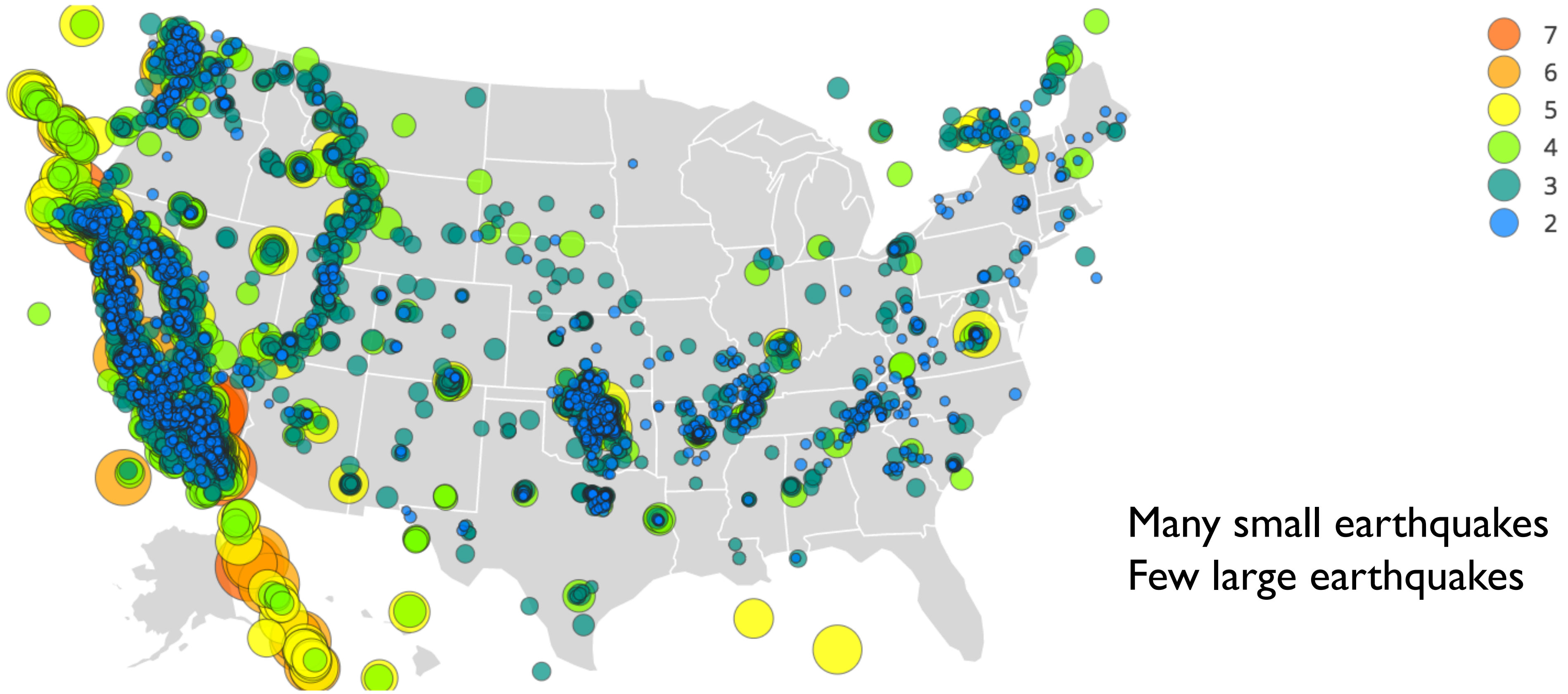
Earthquakes from 1970 to mid 2017



Earthquakes	Did You Feel It Intensity
117, 459	
7287	273,499
4425	85,236
1206	55, 888



EARTHQUAKES WITH DYFI / CDI DATA



Many small earthquakes
Few large earthquakes

DATA IS NOT UNIFORMLY DISTRIBUTED

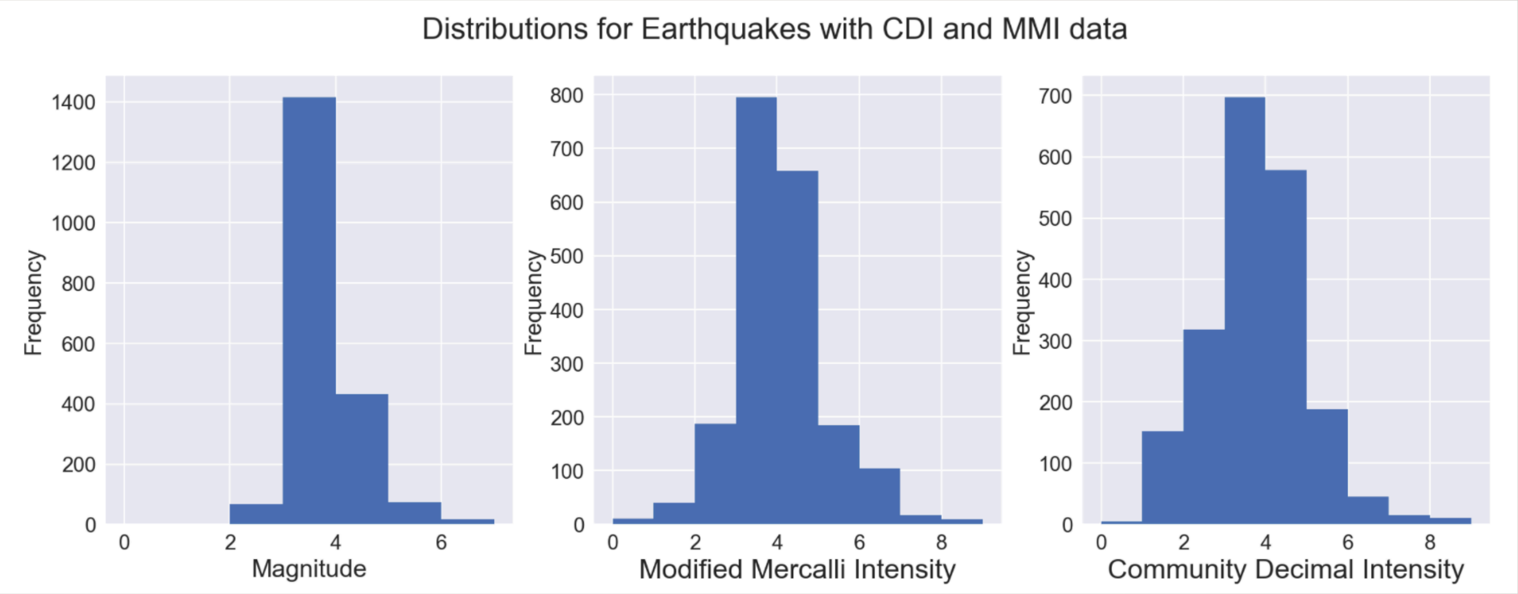
Data from EQ Catalog (not DYFI)

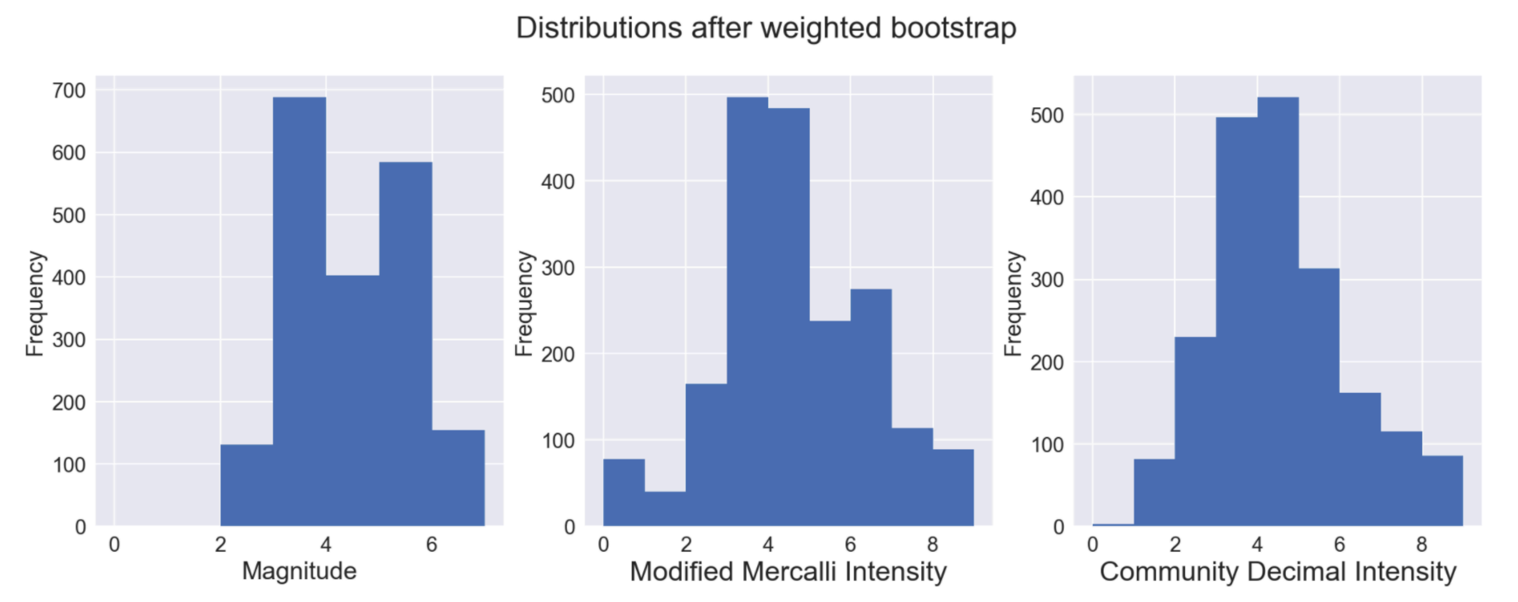
There is a Problem:

- Many small earthquakes
- Few large earthquakes
- MMI not determined for small quakes

And a Bigger Problem:

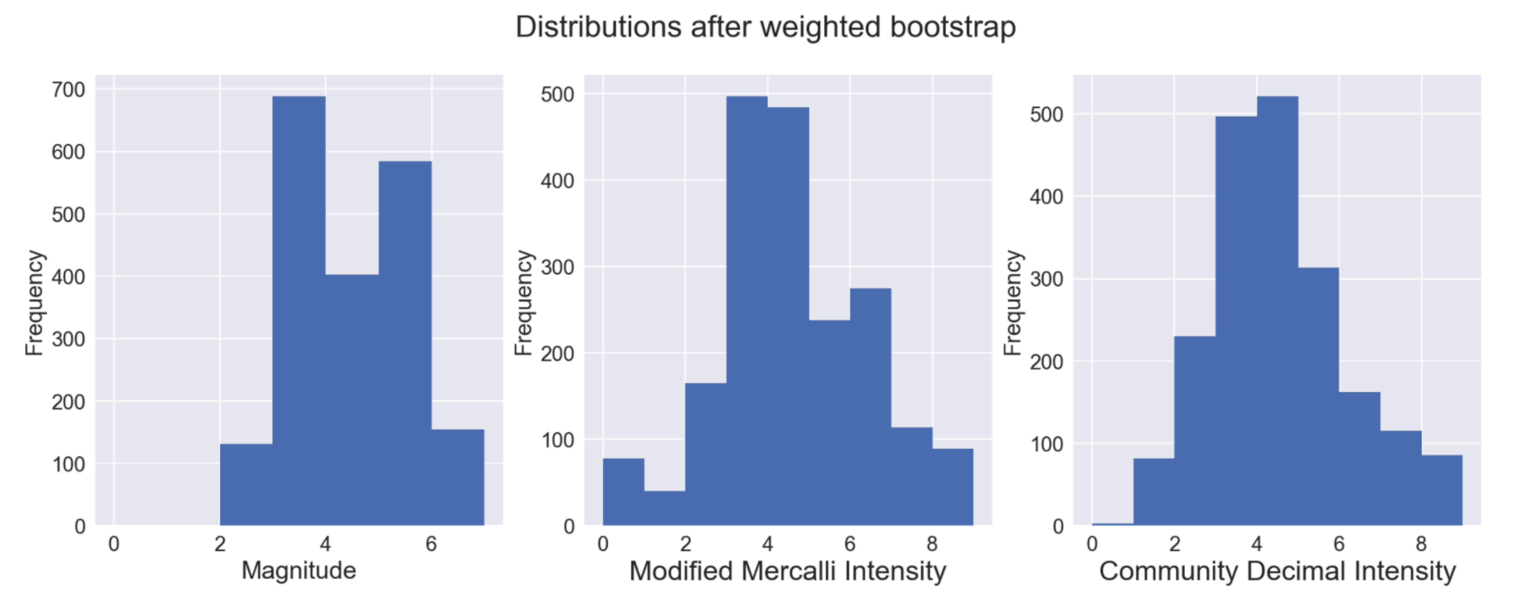
- CDI intensities are seemingly lower in intensity than MMI intensity
- Either reporters are reluctant to report more significant effects, or
- Process for calculating CDI from DYFI reports is flawed

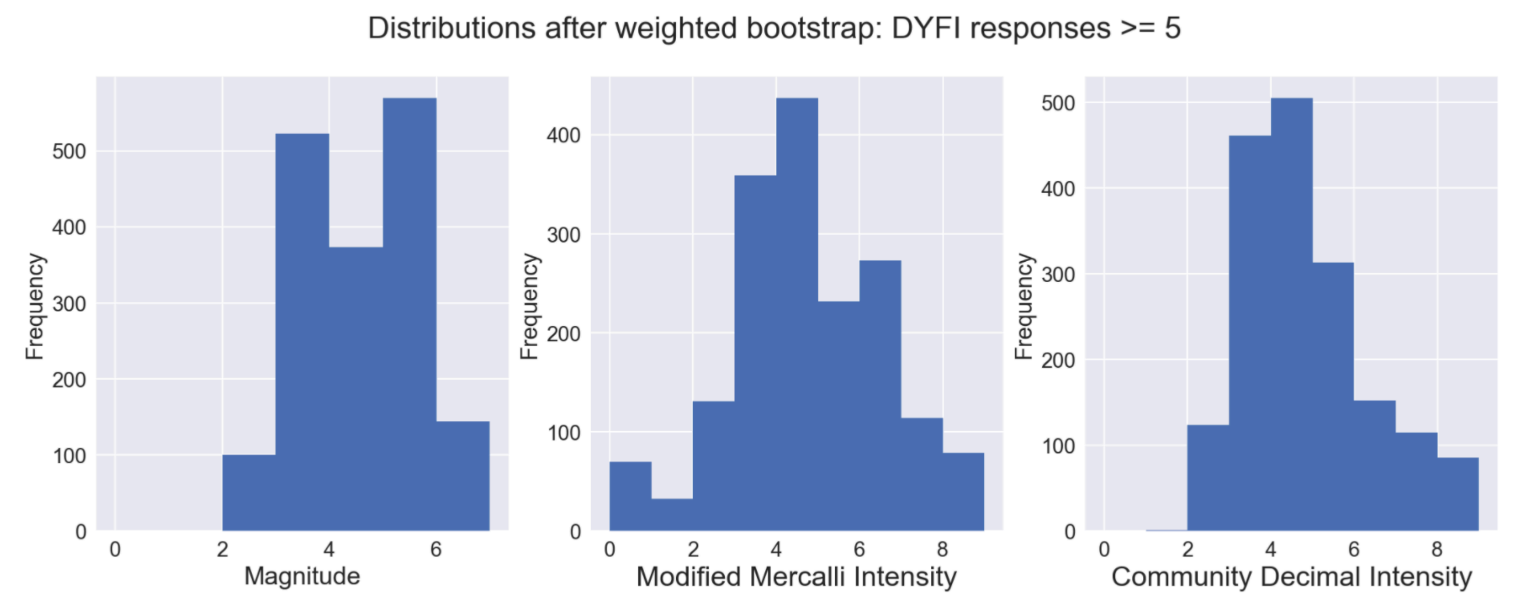




WEIGHTED BOOTSTRAP

- Generate Pseudo Random Number (prn)
- Select next sample
- If prn > category weight, append sample to bootstrap data, else discard (with replace)
- Repeat until new bootstrap is same length as original dataset
- Weights:
  - Mag <= 3 0.75
  - Mag <= 4 0.95
  - Mag <= 5 0.9
  - Mag <= 6 0.1
  - else 0

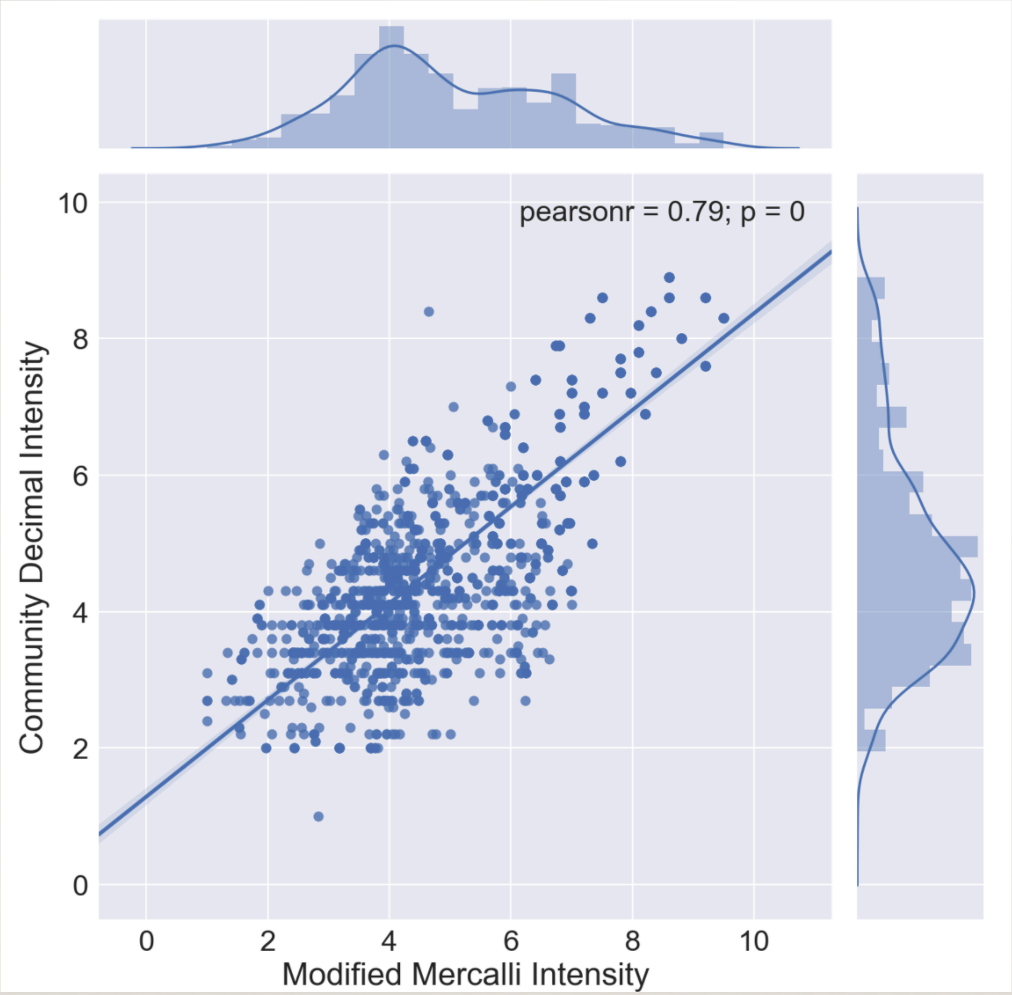


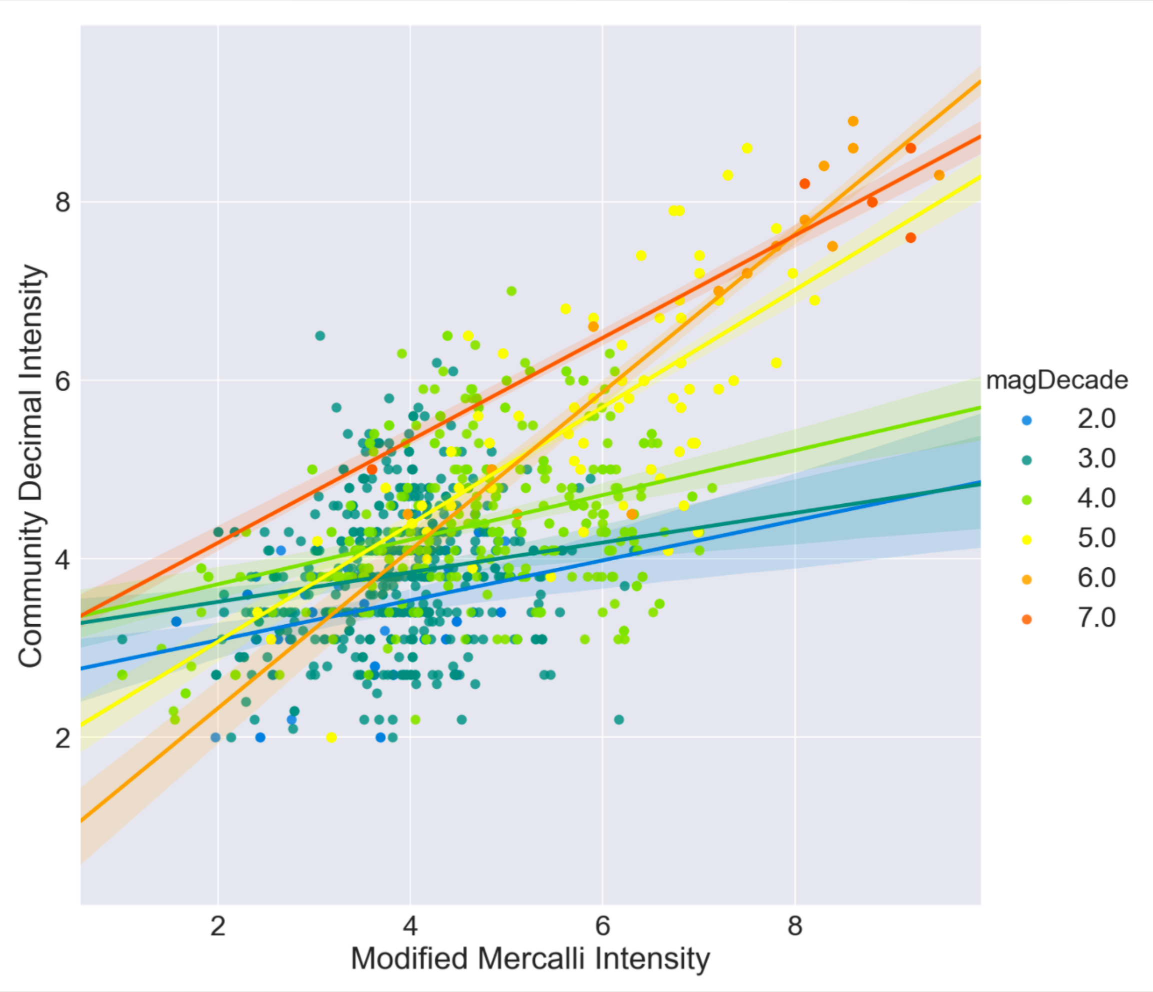


CDI / MMI COMPARISON

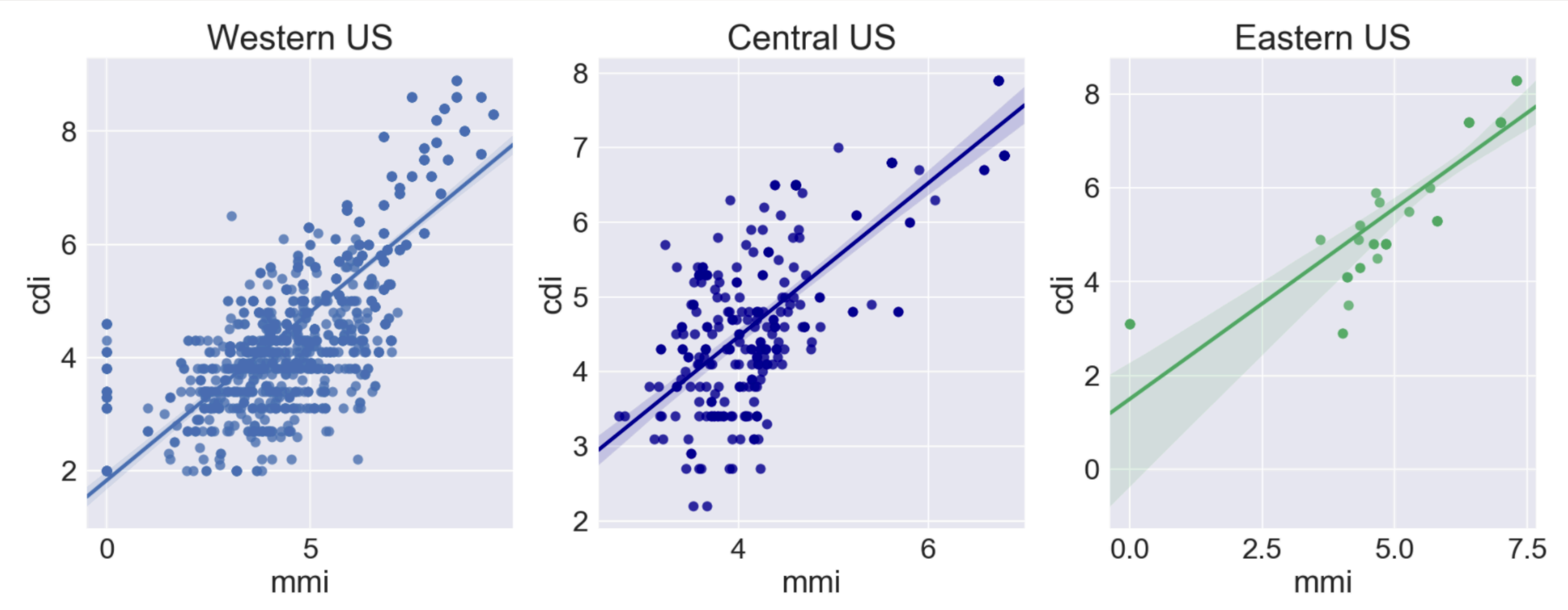
Maximum Intensity Data

Can one predict the other?





REGIONAL DIFFERENCES



PREDICT "EXPERT" MMI FROM "CITIZEN" CDI

Method: Classification Models

- Convert MMI to Integer Numbers
  - 11 prediction classes [0 1 2 3 4 5 6 7 8 9 10]
- Bootstrap with weighting to increase large quakes
- Use additional information:
  - US Region: West, Central, East
  - Depth of earthquake
  - Number of DYFI reports (many are better than one)

Modeling Path

- Logistic Regression
- CV with Bucket of Classifiers (Random Forest best)
- Random Forest alone
- Gridsearch for best Random Forest depth
- Decision Tree to build tree for explanation

- Add more factors and repeat
- Logistic Regression for first pass
- Gridsearch for best strength and penalty type
  - Determine list of contributing factors
- CV with Bucket of Classifiers (Gradient Boost best)
- Gridsearch for best tree depth

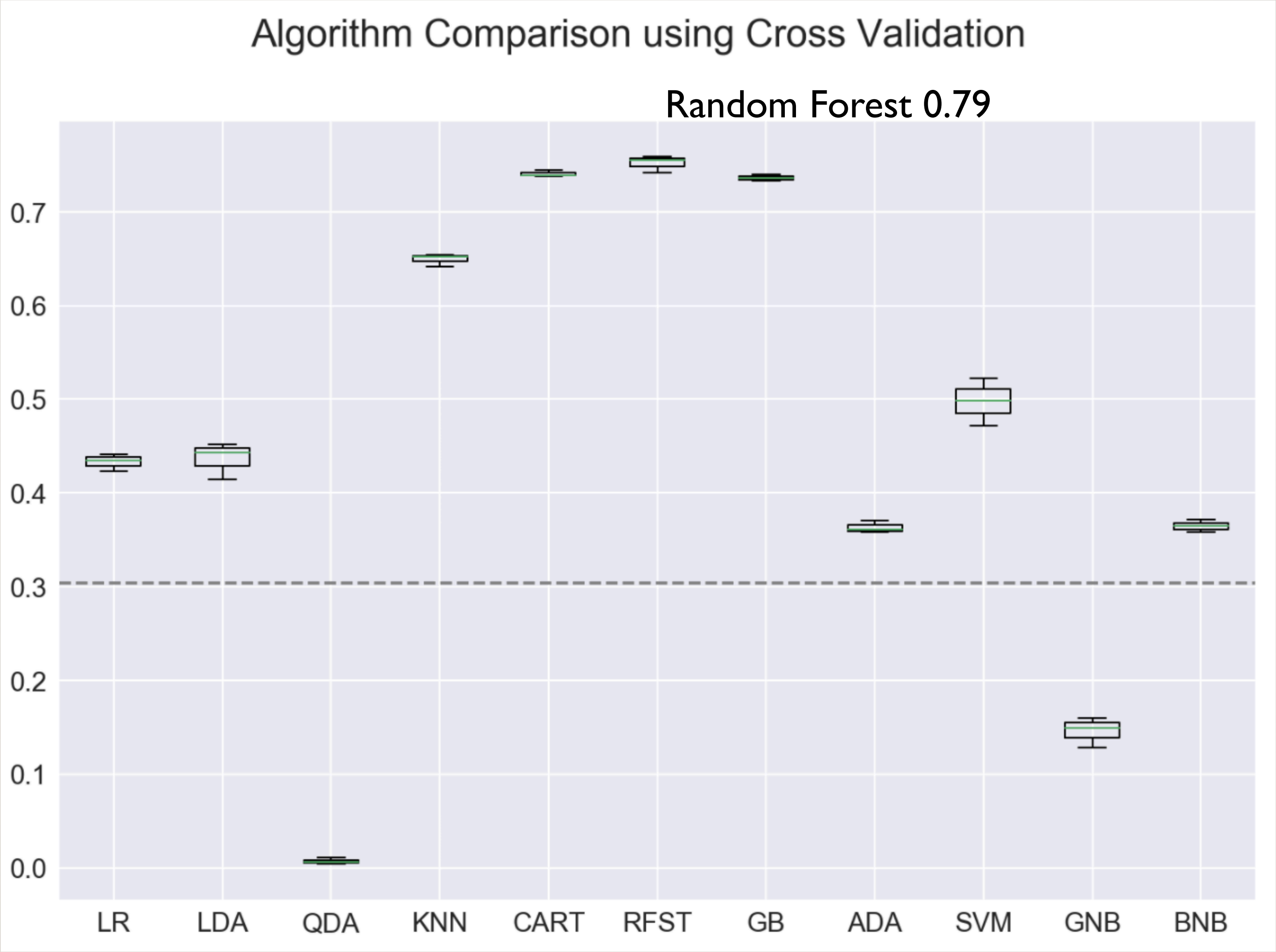
Classification Models	Accuracy Score	Y vs Predicted Pearson R
Logistic Regression on CDI, depth, # CDI reports, US Region, C=1, pen=l2	0.449	0.76
CV with Suite of Classifiers: Random Forest was best	0.789	
Random Forest alone, using same training/test set	0.794	0.94
GridSearch on depth of random forest	0.806	0.93
Decision Tree (run just to get tree graph for explanatory chart)	0.634	
LR with additional features (hour of day, crust thickness and density, reporting station, magnitude calculation details)	0.652	0.91
LR with GridSearch on expanded factors (C=14.51, Lasso)	0.675	0.93
CV with Suite of Classifiers using expanded factors: Gradient Boost	0.854	
Gradient Boost alone, using expanded factors and same training/test set		

PREDICT MMI FROM CDI

BASIC FACTORS

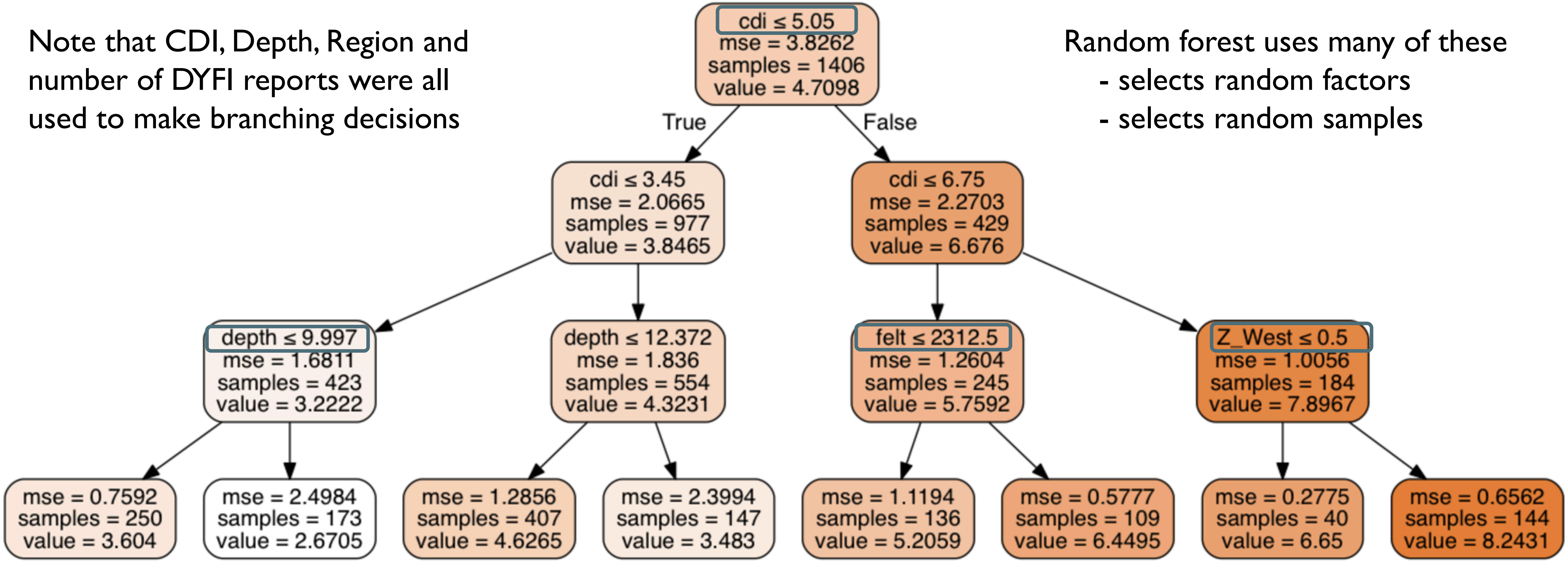
Classifiers Employed

- LR: LogisticRegression()
- LDA: LinearDiscriminantAnalysis()
- QDA: QuadraticDiscriminantAnalysis()
- KNN: KNeighborsClassifier()
- CART: DecisionTreeClassifier()
- RFST: RandomForestClassifier()
- GB: GradientBoostingClassifier()
- ADA: AdaBoostClassifier()
- SVM: SVC()
- GNB: GaussianNB()
- BNB: BernoulliNB()



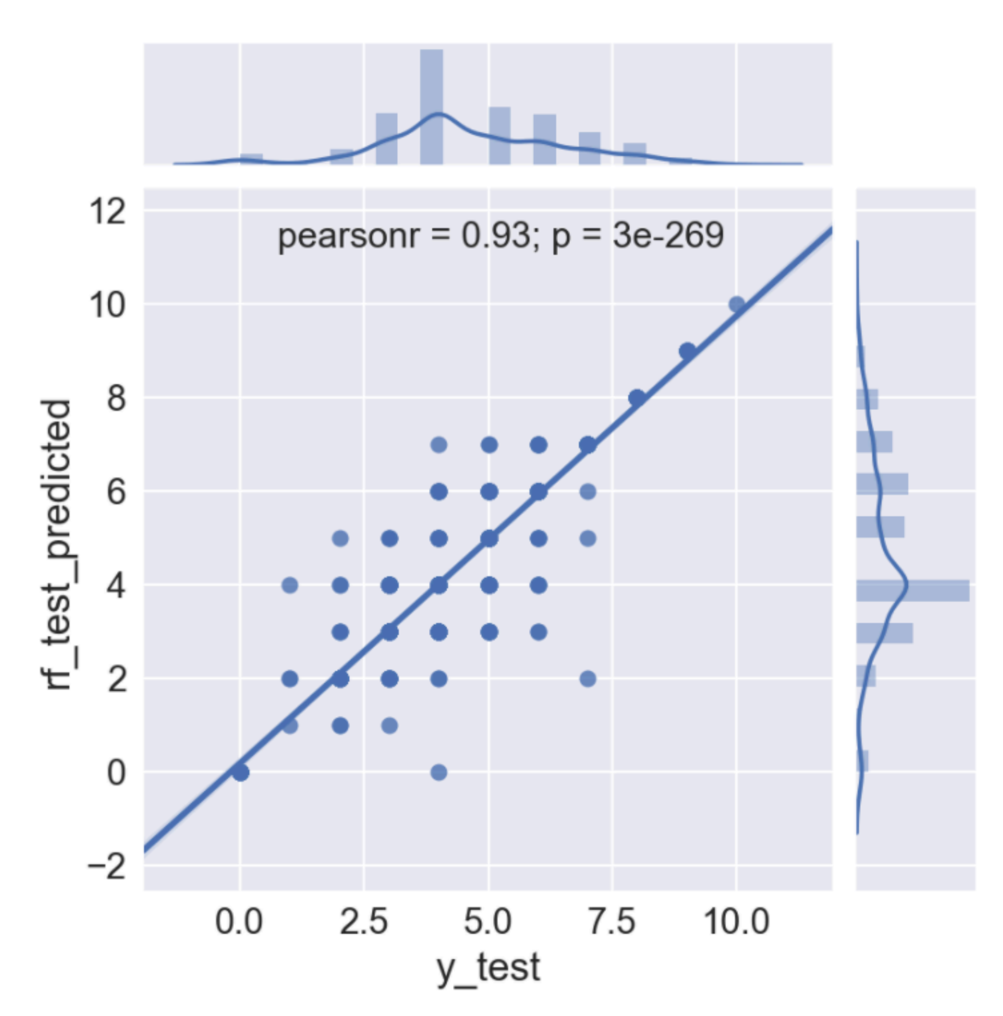
DECISION TREE SELECTION PROCESS

Note that CDI, Depth, Region and number of DYFI reports were all used to make branching decisions



Random forest uses many of these
- selects random factors
- selects random samples

PREDICT "EXPERT" MMI FROM "CITIZEN" CDI



Method: Random Forest

Convert MMI to Integer Numbers

Use additional information:

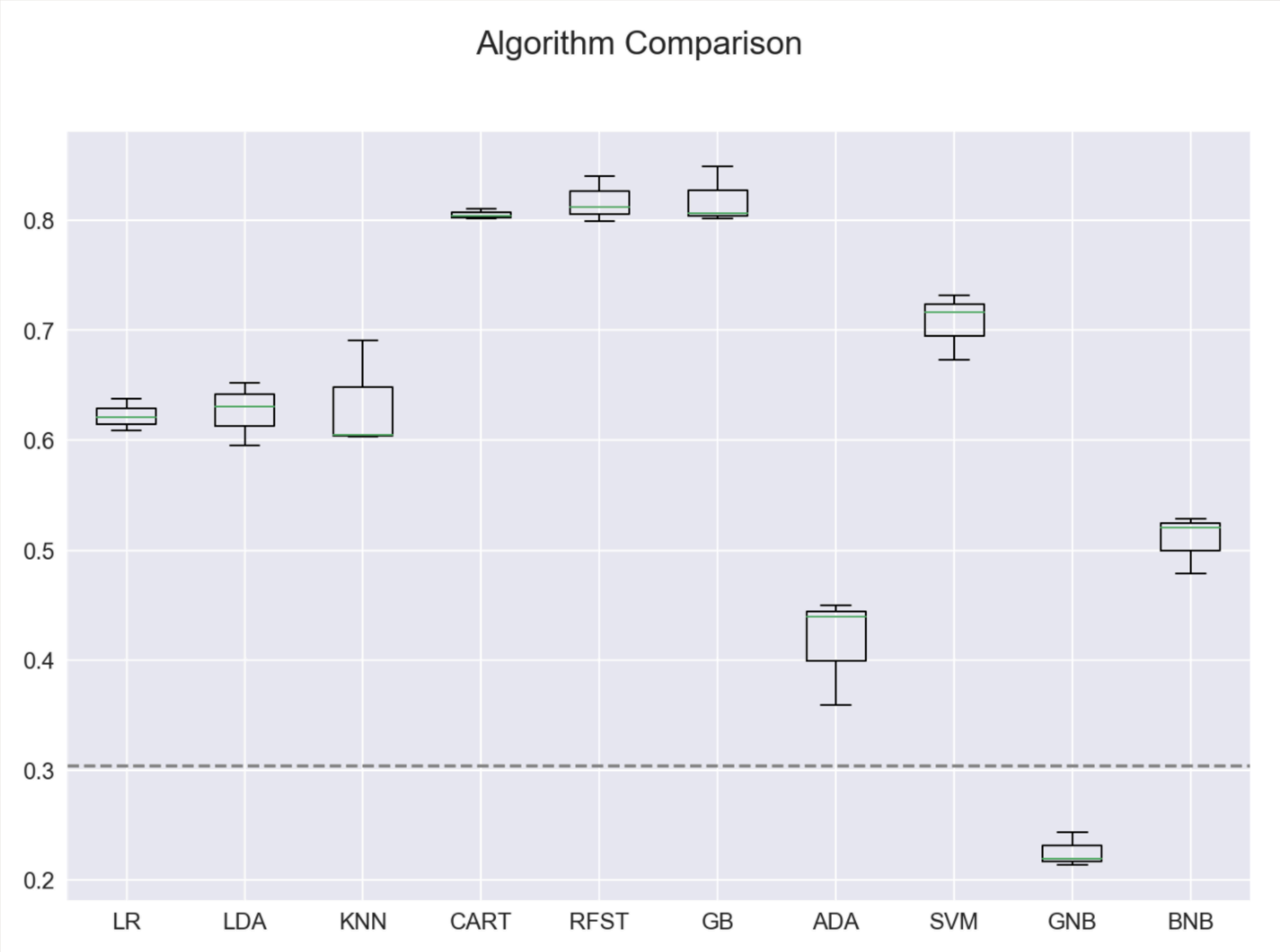
- US Region: West, Central, East
- Depth of earthquake
- Number of DYFI reports (many are better than one)

PREDICT MMI FROM CDI

EXPANDED FACTORS

Classifiers Employed

- LR: LogisticRegression()
- LDA: LinearDiscriminantAnalysis()
- KNN: KNeighborsClassifier()
- CART: DecisionTreeClassifier()
- RFST: RandomForestClassifier()
- GB: GradientBoostingClassifier()
- ADA: AdaBoostClassifier()
- SVM: SVC()
- GNB: GaussianNB()
- BNB: BernoulliNB()



PREDICT INTENSITY WITH DISTANCE

Model: Multiple Linear Regression
with Regularization

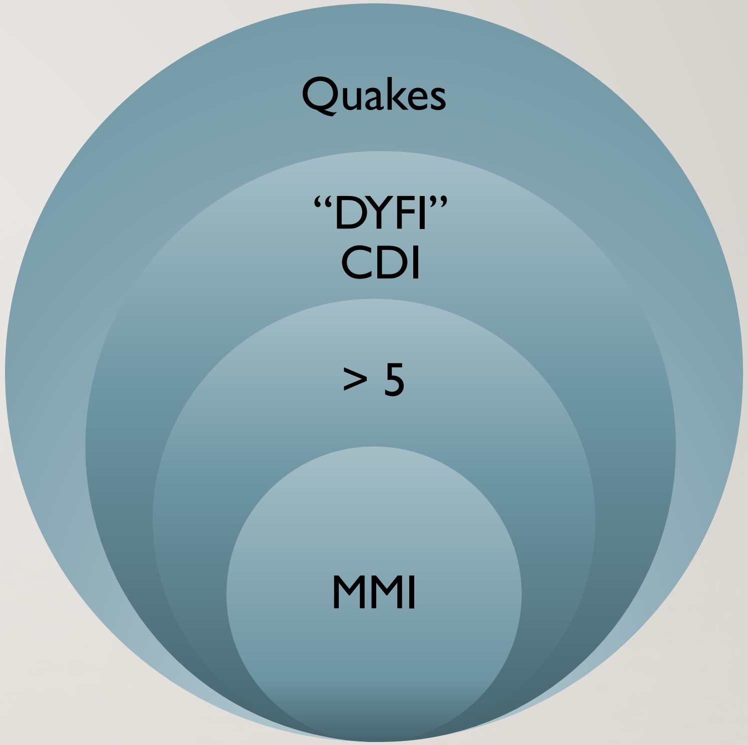
Data: 1206 Quakes with 55,888 DYFI ”zip codes”

Method:

- Predict Community Decimal Intensity (CDI)
- Predict Modified Mercalli Intensity (MMI)
- Compare with published decay equations

Use traditional factors:
Distance, Magnitude
Number of DYFI reports

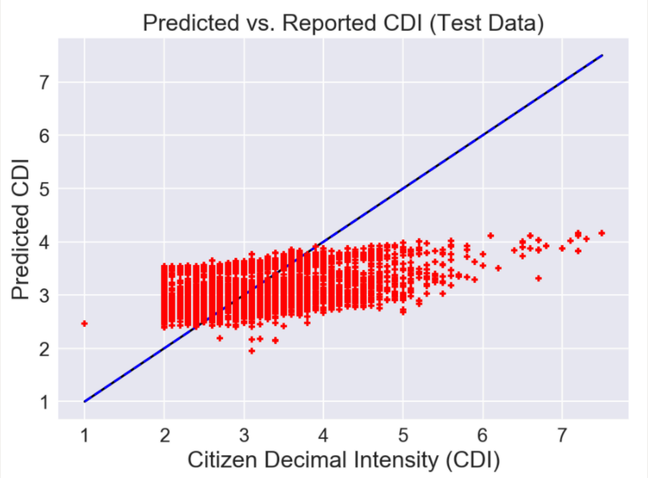
Use expanded factors:
Distance, Magnitude, Depth, Crust Parameters,
Magnitude Calculation Method

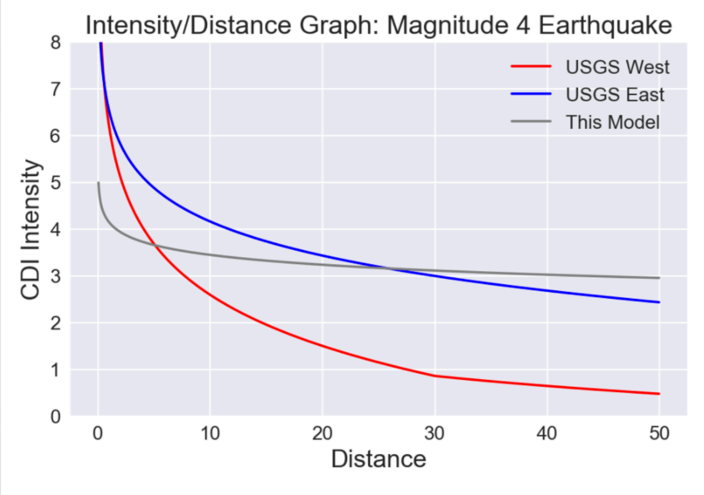


PREDICT CDI AT DISTANCE

R^2: 18% - difference between the average CDI earthquake intensity for all observations is explained by modeling with:

Magnitude & Distance only

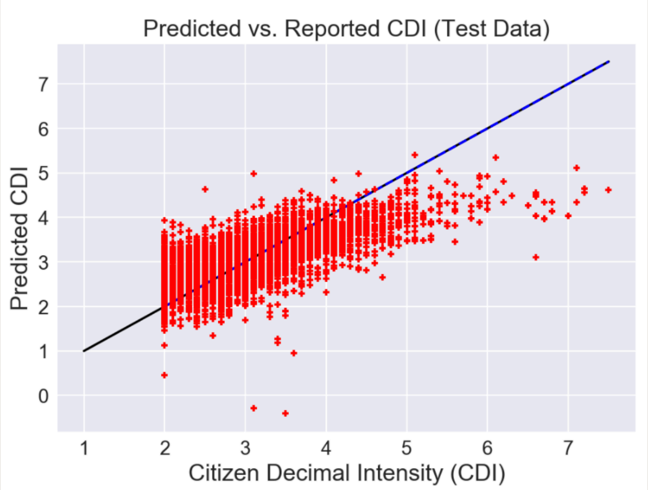


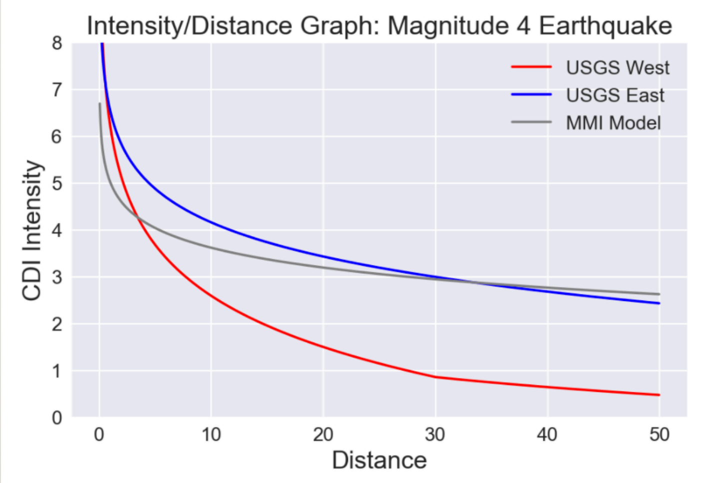


R^2: 45% - difference between average CDI intensity of all earthquakes and what is explained by modeling with:

Magnitude and Distance   and …

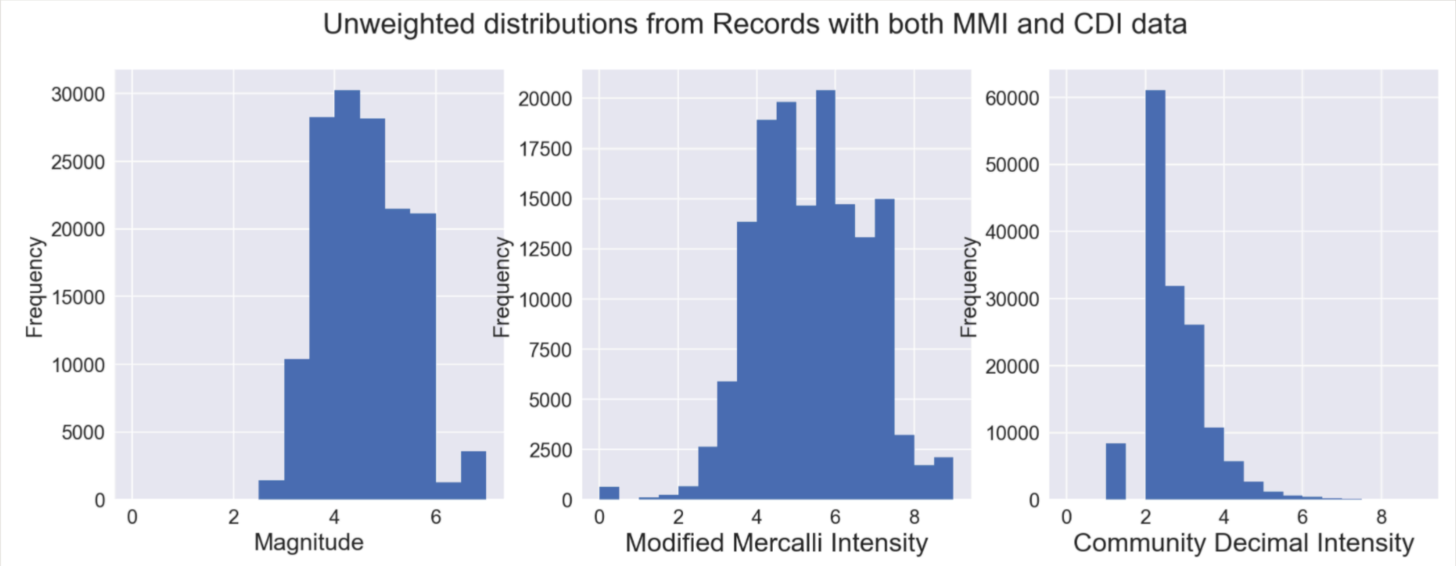
- Depth,
- Crustal Density & Thickness
- Method of computing magnitude

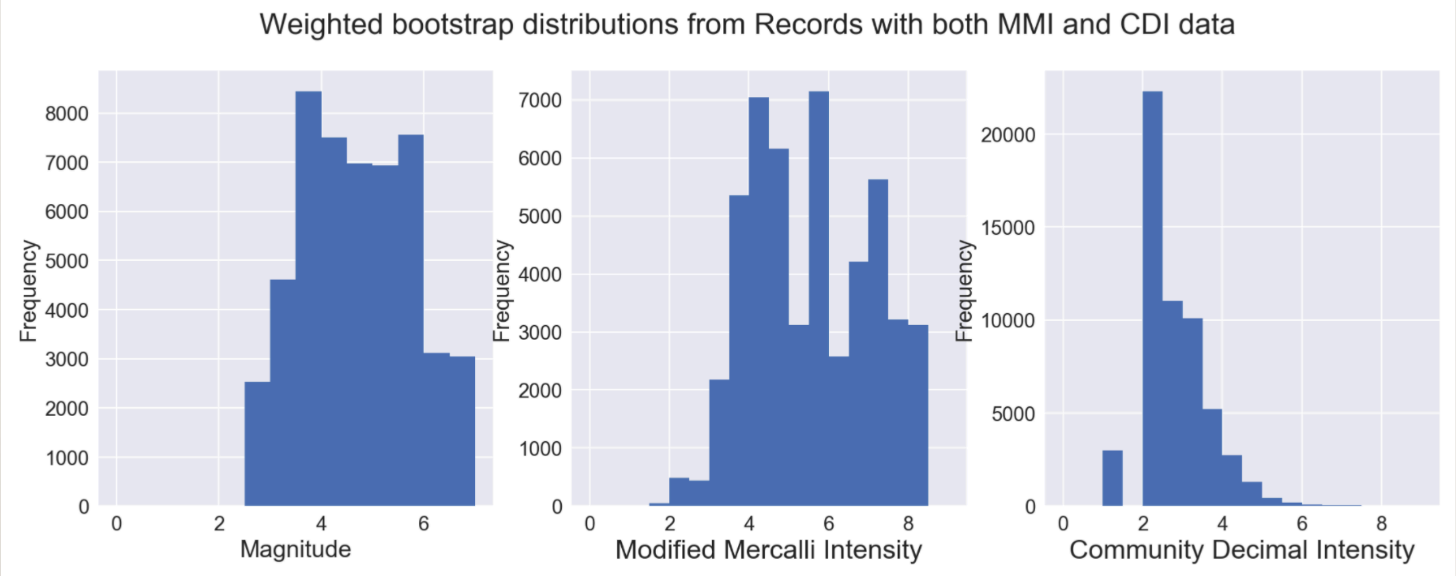




DISTRIBUTION CHARACTERISTICS OF SAMPLE DATA

Filtered to samples with
- CDI
- MMI

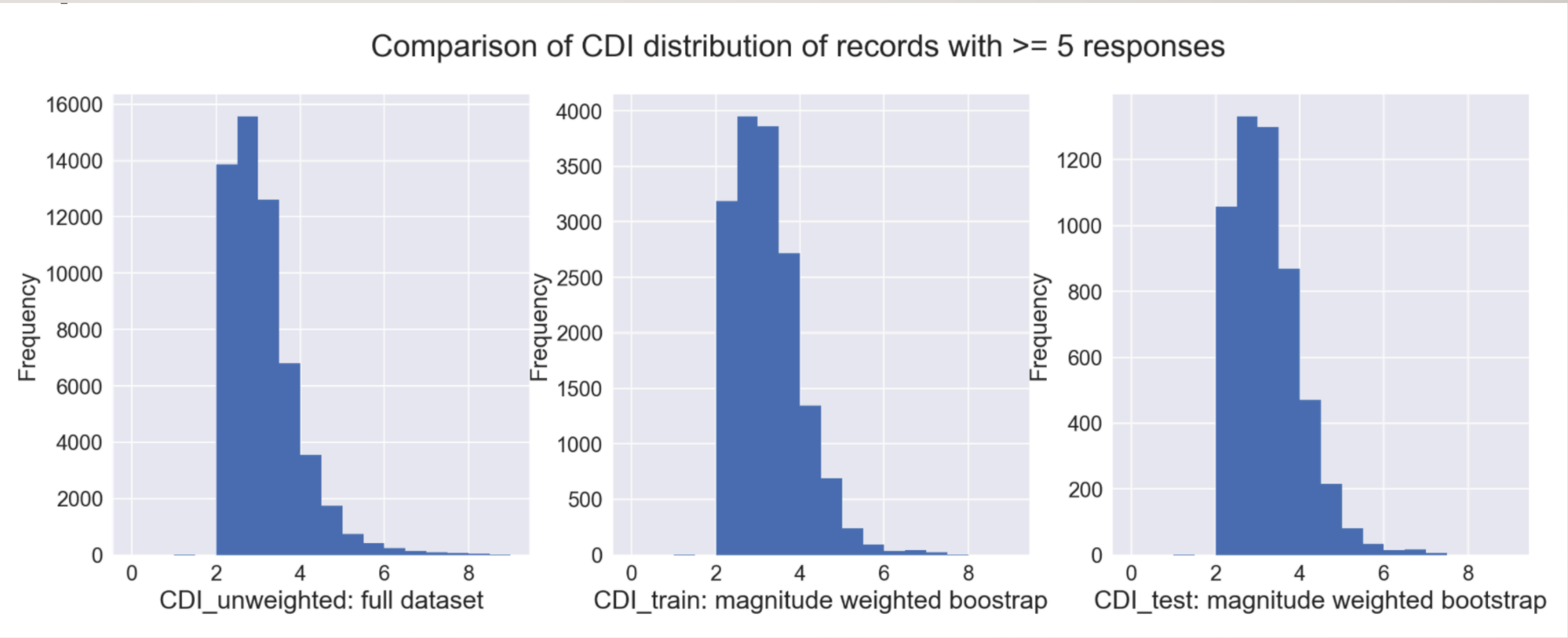




DISTRIBUTION CHARACTERISTICS CDI TEST/TRAIN

Weighted Bootstrap filtered to samples with

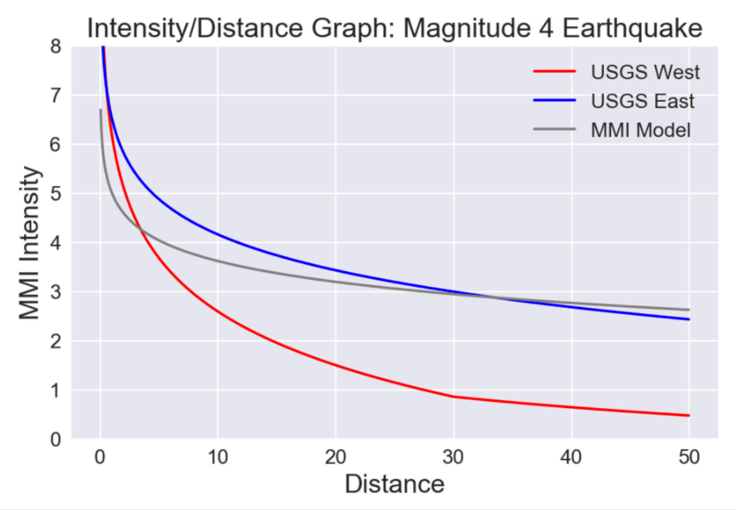
- CDI (with 5 or more records per quake)
- MMI

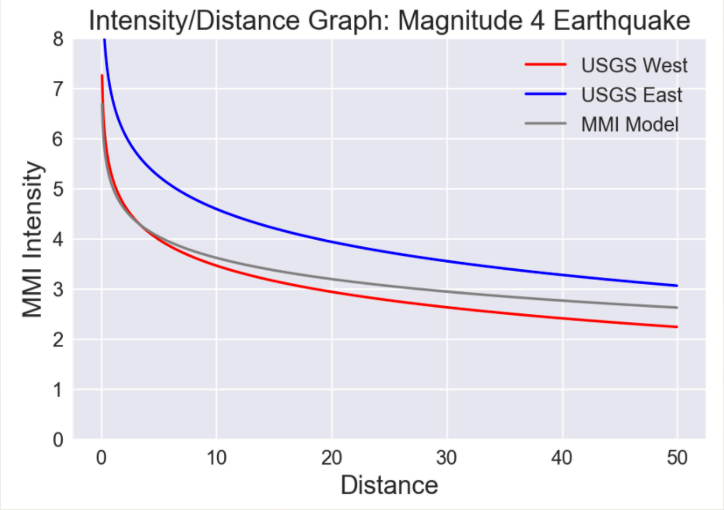


COMPARISON WITH USGS MMI CURVE

Interestingly, this model's predictions align more closely with the USGS equations for predicting MMI

Possibly because additional factors draw CDI toward MMI scale?





PREDICT MMI FROM MAGNITUDE AND DISTANCE

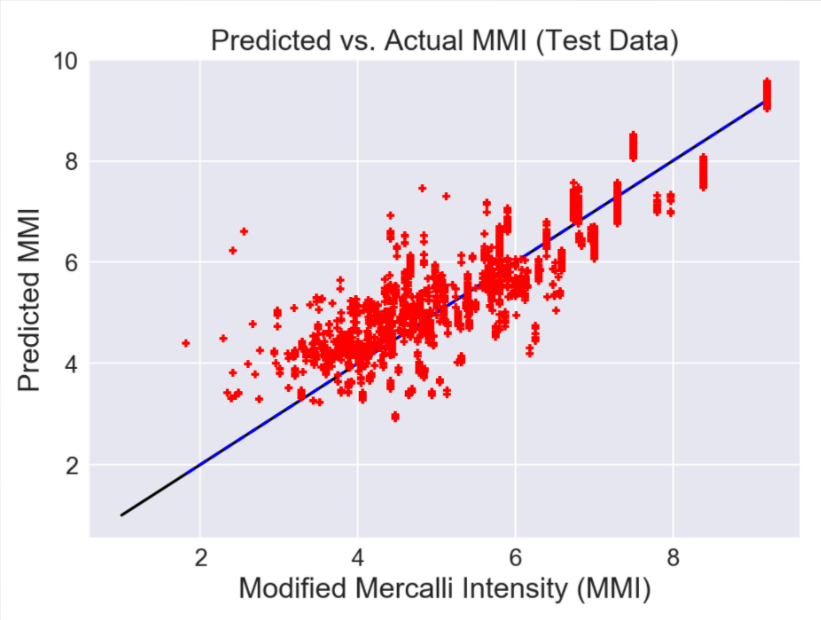
Modified Mercalli Intensity (MMI)

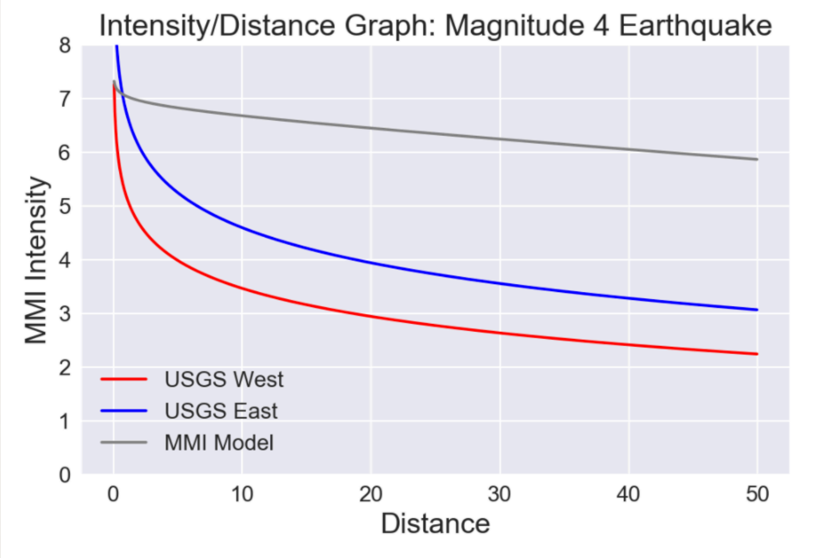
This explains 91% of the difference between average MMI intensity of all earthquakes and actual intensity observed –

… and only magnitude and distance were used!

What's up with this?

- Data reported and reviewed by experts
- Reported over longer period of time
  - Includes larger proportion of high magnitude earthquakes

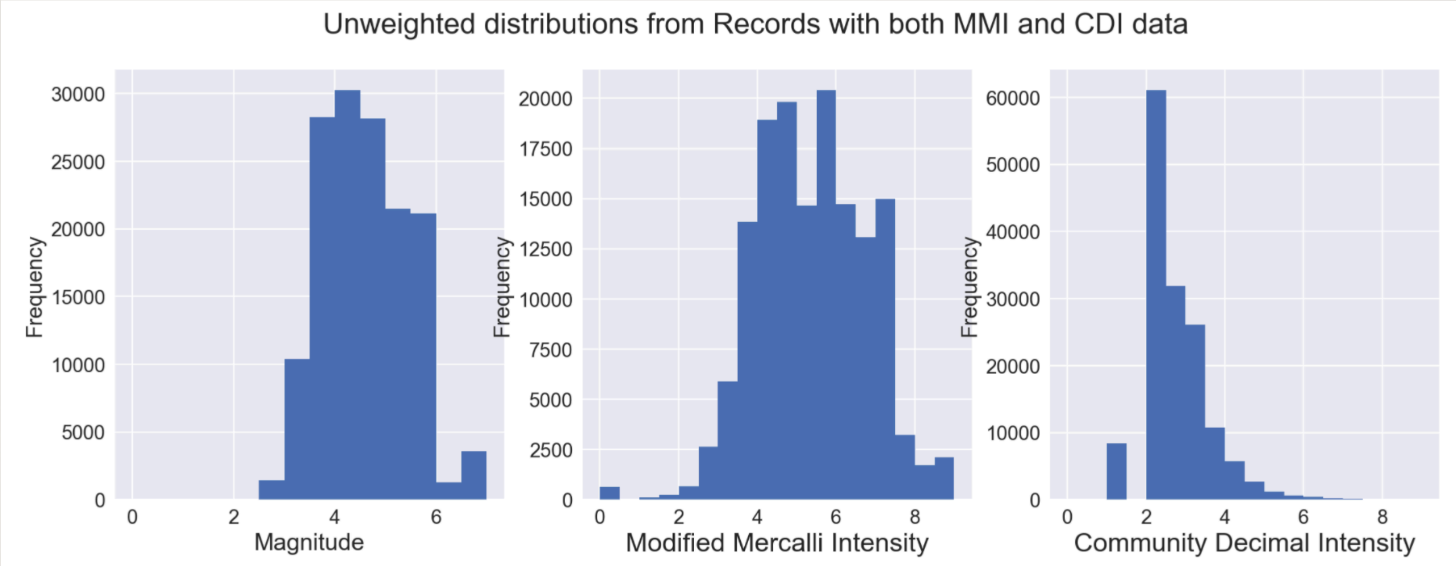


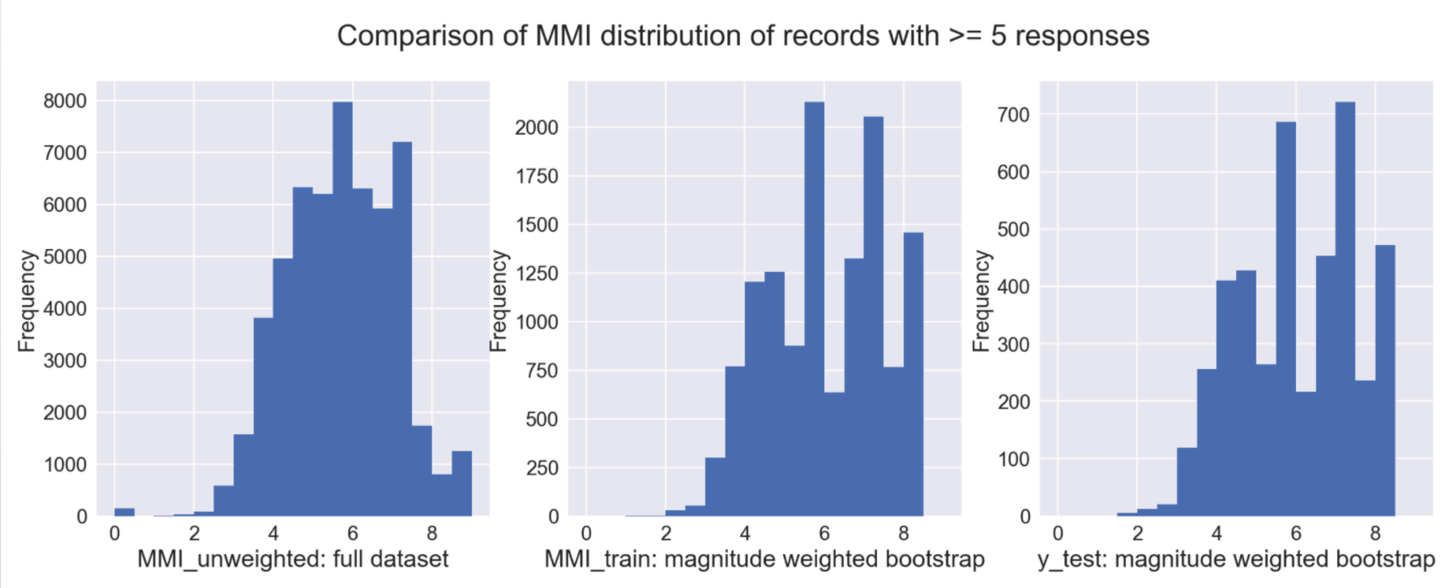


DISTRIBUTION CHARACTERISTICS OF SAMPLE DATA

Filtered to samples with

- CDI (with 5 or more records per quake)
- MMI





DATA SOURCES

- USGS: Common Catalog
  - Aggregation of Earthquake Events from many reporting agencies
  - Includes "Did You Feel It," CDI data, MMI and Magnitude
- LLNL: Crust 1.0 Model
  - Crustal Boundary and Density Data

References

- Wald, David J. (2011). USGS "Did You Feel It?" Internet-based macroseismic intensity maps. ANNALS OF GEOPHYSICS, 54.6
- Oklahoma Example: Matthew Schroyer paper
- Crowdsourced Geographic Information use in Govt.

GLOBAL CRUSTAL MODEL

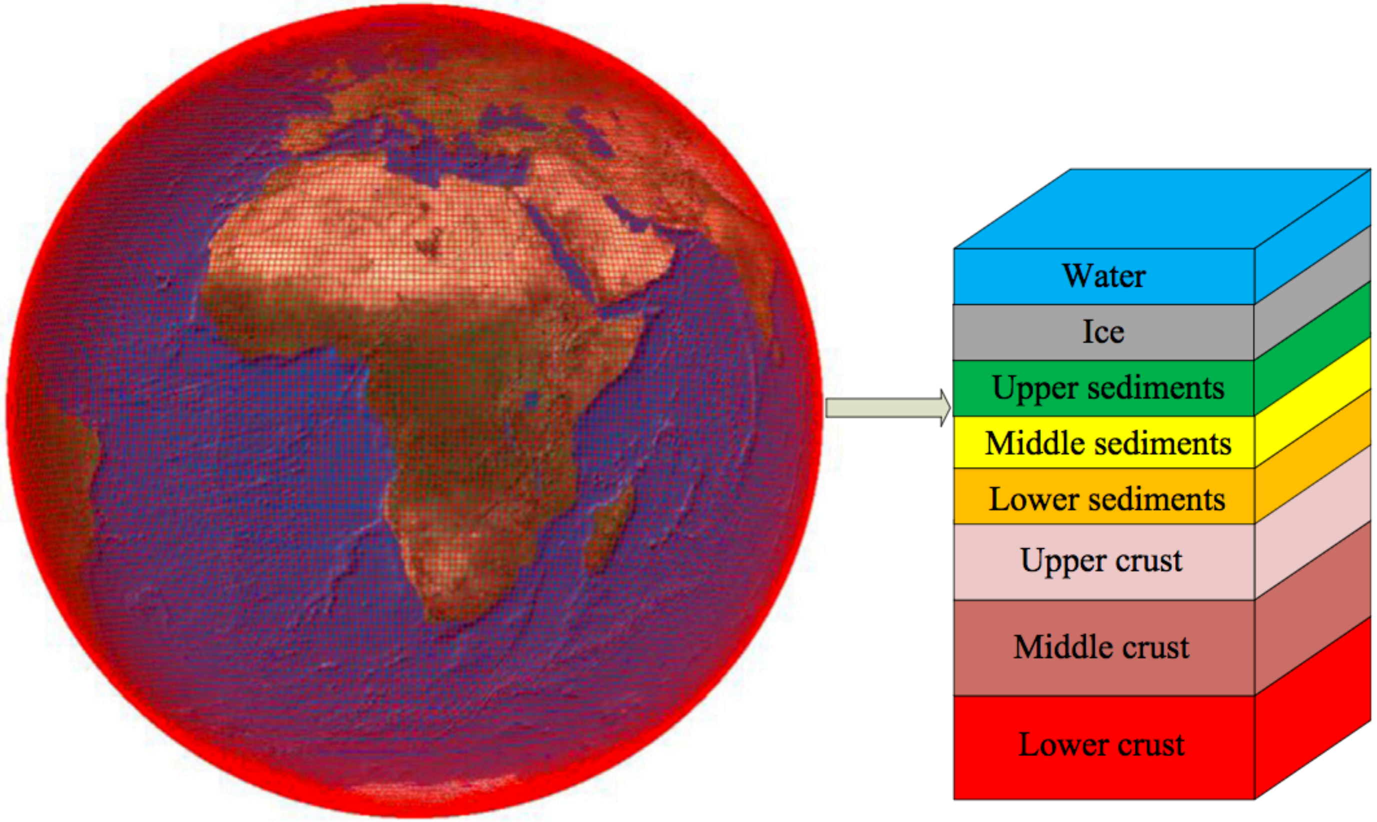


Fig. 1. Crustal structure representation implemented in CRUST 1.0. The crust is parametrized laterally by 64800 $1^\circ \times 1^\circ$ latitude-longitude grids and vertically as eight geophysically identified sublayers.

Zhu, L.F., Pan, Z., Sun, J.Z., (2016). Visualization and dissemination of global crustal models on virtual globes, Computers and Geosciences v90, Part A, 34-40. Elsevier, https://doi.org/10.1016/j.cageo.2016.01.015

https://igppweb.ucsd.edu/~gabi/crust1.html#visualization

Intensity	Shaking
I	Not felt
II	Weak
III	Weak
IV	Light
V	Moderate
VI	Strong
VII	Very strong
VIII	Severe
IX	Violent
X	Extreme

DID YOU FEEL IT?

A decision tree is used to calculate CDI from these reports. CDI = Community Decimal Intensity

https://earthquake.usgs.gov/earthquakes/eventpage/unknown#tellus

Felt Report – Tell Us!

OMB No. 1028-0048
Expires 05/31/2018

Your location when the earthquake occurred

Choose Location

Time of Earthquake **Local time *1/31/2008 9:00 AM*, or Relative time *5 minutes ago***

Did you feel it?

- Yes
- No

The remainder of this form is optional.
Help make a shaking intensity map by telling us about the shaking at your location.

What was your situation during the earthquake?

- Not specified
- Inside a building
- Outside a building
- In a stopped vehicle
- In a moving vehicle
- Other

Please describe

Were you asleep?

- Not specified
- No
- Slept through it
- Woke up

Did others nearby feel it?

- Not specified
- No others felt it
- Some felt it, most did not
- Most felt it
- Everyone/almost everyone felt it

How would you describe the shaking?

- Not specified
- Not felt
- Weak
- Mild
- Moderate
- Strong
- Violent