Modeling the Future Challenge

Anan Abrar, Danyil Blyschak, Jeremy Kim, Hiren Kumawat

Executive Summary

Farming is largely contingent on luck and the known unknowns. Farmers choose to cultivate a crop because they expect the market price of it to be greater than the costs required to cultivate it. As supply goes up while demand stays constant, price decreases. In order to prevent farmers from going bankrupt due to overproducing, farming subsidies were introduced. Additionally, subsidies attempted to address the uncertainty associated with natural causes of loss, be it drought or a hurricane, by providing a buffer by paying per unit of crops.

Currently, there are two forms of subsidies. PLC program payments pay farmers when the market price of their crops drop below a certain amount. As a result, it covers farmers from ruining themselves by placing a buffer on the detriments of overproduction. ARC program payments pay farmers when the revenue of their farm drops below a certain amount determined by historical data concerning the nearby counties. Thus, ARC not only protects farmers from a bad market price, but environmental losses and significant agricultural disasters as their revenue would be low as a result.

Despite these subsidies, (small) farmers are often going bankrupt. As a means of counteracting that, as well as providing them incentives to adopt innovative technology to minimize their acreage loss and maximize its productivity, we propose three solutions. First, introducing a rebate program no different from when solar panels are implemented would distort incentives and encourage farmers to invest in the latest technology. Second, insurance companies would compete amongst each other as the new technology ensures that the farmers are safer and less prone to claiming insurance losses. Finally, providing new subsidies surrounding rotation of crops ensures long term soil health. The three

aforementioned solutions should mitigate climate change's impact on crop losses, having positive externalities such as less water usage and greater land efficiency.

Our analysis began with the selection of two states. We believed that the two states should fundamentally differ in several factors in order to make our modeling more applicable to other regions. Thus, we chose the states of California and Louisiana since the crops that grow in these two states are similar, but the level of agricultural technology that California has implemented is higher than that of Louisiana. Additionally, we decided to study the rice output in these states because of the recent introduction of new agriculture technologies such as seed varieties and precision farming.[1]

It should be noted that we analyzed nationwide data simply because specific datasets of parameters we were interested in were more abundantly available for states than they were for the nation as a whole.

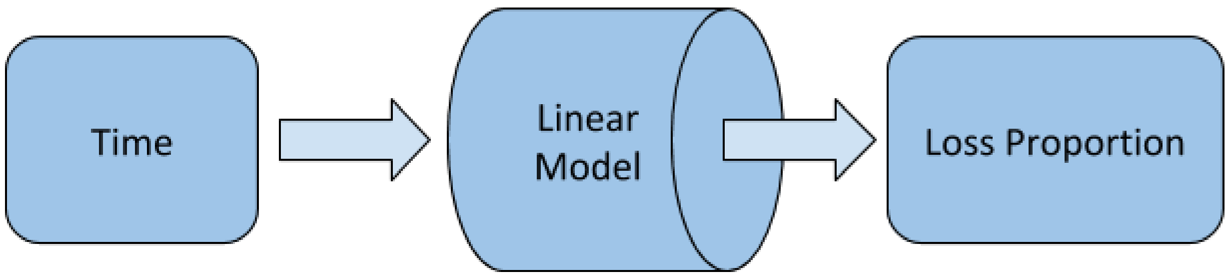
Our modeling goal was to predict the projected loss for a farm given a certain set of conditions. However, instead of predicting the indemnities, we changed the target to a new index we propose called the loss proportion: acres claimed as losses / acres harvested that year.

Now that we identified our research topic, we needed to engineer features of our model. These features would serve as inputs for our regression model to predict loss proportion as a function of time. We started with a basic model: $f(\widehat{t}) = \widehat{Y}$. $\widehat{t}$ is a list of years and $\widehat{Y}$ is a list of

[1] McBride, W., Skorbiansky, S. R., & Childs, N. (2018). U.S. Rice Production in the New Millennium: Changes in Structure, Practices, and Costs. *Economic Information Bulletin*. Retrieved from https://www.ers.usda.gov/webdocs/publications/90926/eib-202.pdf?v=4597.3

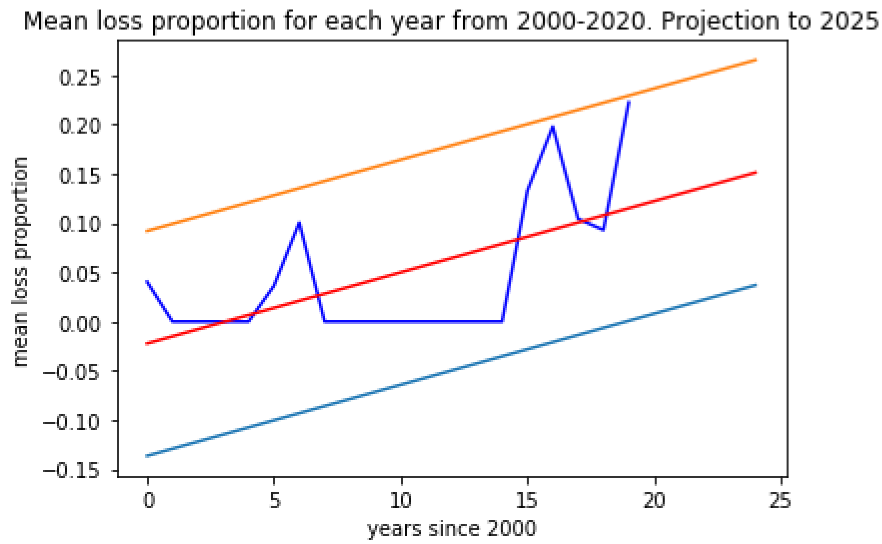
each year's mean loss proportion per farmer. We call this our naive model because it does not take into account any factors concerning crop growth in Louisiana. A visual depiction of our regression model can be seen below in **Figure 1**.

Figure 1: Flowchart for naive linear model



To develop this model, we filtered our pre-processed USDA primary dataset so only the data on Louisiana counties that grew rice remained in the spreadsheet. We were left with 7832 data points on farmers' loss proportions spread across 30 counties in Louisiana. We calculated the average expected loss proportion for each year from 2000-2019, a vector of size 20x1, and initialized a linear regression model using Python's Sci-Kit Learn machine learning library. Since our target is a vector of size 20x1, this model's input vector is also size 20x1, where each element of the input is the value of the year we wish to determine the mean loss proportion for. We ran the linear regression and got unsatisfactory results. As seen in **Figure 2**, Our r^2 value was 0.36 and our 95% confidence interval was too wide, indicating that the model had little knowledge about the data's trend.

Figure 2: Initial Results of Our Naive Regression



We took a step back to ask ourselves what we were trying to accomplish by using time as a feature. We realized we used time in an attempt to inform the model on temperature trends. A 2020 study from the National Oceanic and Atmospheric Administration (NOAA) found that global surface temperatures were rising every year.[2] Keeping this study in mind, we thought that if we gave the model a certain year, the model would correlate increasing time with increasing surface temperatures, which would increase loss proportions. We realized that our fit of the mean annual loss proportions was poor because time could be also correlated with a host of unrelated factors.

As a result of our naive model's results, we realized that we had to be more explicit with identifying trends we wanted our model to incorporate when it projected future loss proportions. The Environmental Protection Agency[3] claimed that agricultural productivity was heavily

---

[2]Lindsey, R., & Dahlman, L. (2020, January 16). Climate Change: Global Temperature. Retrieved February 28, 2020, from Climate website: http://Climate Change: Global Temperature

[3] Climate Impacts on Agriculture and Food Supply. (n.d.). Retrieved February 29, 2020, from EPA website: https://19january2017snapshot.epa.gov/climate-impacts/

dependent on climate so we used incorporated the NOAA climate dataset into our existing dataset. We extracted Louisiana's monthly average temperature, maximum temperature, minimum temperature, precipitation, cooling degree days, heating degree days, Palmer Drought Severity Index, Palmer Hydrological Drought Index, Palmer Modified Drought Index, and the Palmer Z-Index. All of these variables are relevant to describing Louisiana's climate and access to resources like water. In addition to the values associated with each variable, we also incorporated the variables' departure from the variable's mean value from 1895 to 2020. This departure was called the anomaly and gives the model context for what abnormal conditions would look like (high departure from the mean). Use of these weather parameters is a robust way to approach this regression because our model is both more interpretable.

Now that we introduced these variables, we had an additional task on top of predicting the loss proportion: projecting the NOAA weather variables. Since our new regression model will use those variables, we need to be able to tell the model what features like average temperature will look like 10 years in the future in order for it to make a projection of the average loss proportion in 10 years.

In order to project the NOAA variables (16 in total including the each variable's anomaly), we used a time-series regression model that Facebook released called Prophet. The traditional approach to time-series regression modeling is to use algorithms like ARIMA (autoregressive integrated moving average), which calculates a moving average of values to predict. ARIMA's predictive capabilities are represented with this equation:

$$y_t = \mu + \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} .$$

climate-impacts-agriculture-and-food-supply_.html

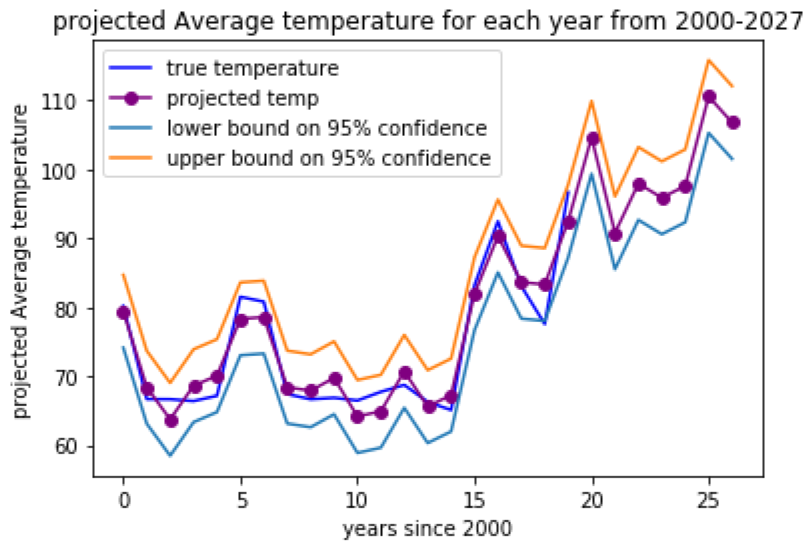
However, there are a few issues with the ARIMA algorithm. First, it requires that measurements are regularly spaced. This spacing is critical to ARIMA creating a moving average but makes a mathematical model hard to scale because we will need to continually maintain the data collection methods. Second, ARIMA models are known to cause large trend errors. Given that our loss proportion prediction model relies on a robust projection of the weather variables, these trend errors would be unacceptable.

Prophet is called a decomposable time series model in that its predictions are based off of three other functions: $y(t) = g(t) + s(t) + h(t) + \varepsilon$. As per Facebook's paper, "*g(t)* is the trend function which models non-periodic changes in the value of the time series, *s(t)* represents periodic changes (e.g., weekly and yearly seasonality), and *h(t)* represents the effects of holidays which occur on potentially irregular schedules over one or more days."[4] This decomposable time series model was an attractive choice for its flexibility and robustness. We ran Prophet on the annual averages for all 16 parameters from 2000 to 2020. As seen in **Figure 3**, we received high r^2 scores, such as 0.92 when projecting the average temperature to 2027.

The key parameter to tune when training Prophet was the trend flexibility. This parameter represents the strength of the sparse prior used to optimize the time series regression. **Figure 4** demonstrates how we found that the optimal flexibility was 0.8, which gives the regression enough degrees of freedom to fit the NOAA training data but not enough to overfit.

[4] Taylor, S. J., & Letham, B. (2017). Forecasting at Scale. *Practical Data Science for Stats*. Retrieved from https://peerj.com/preprints/3190/

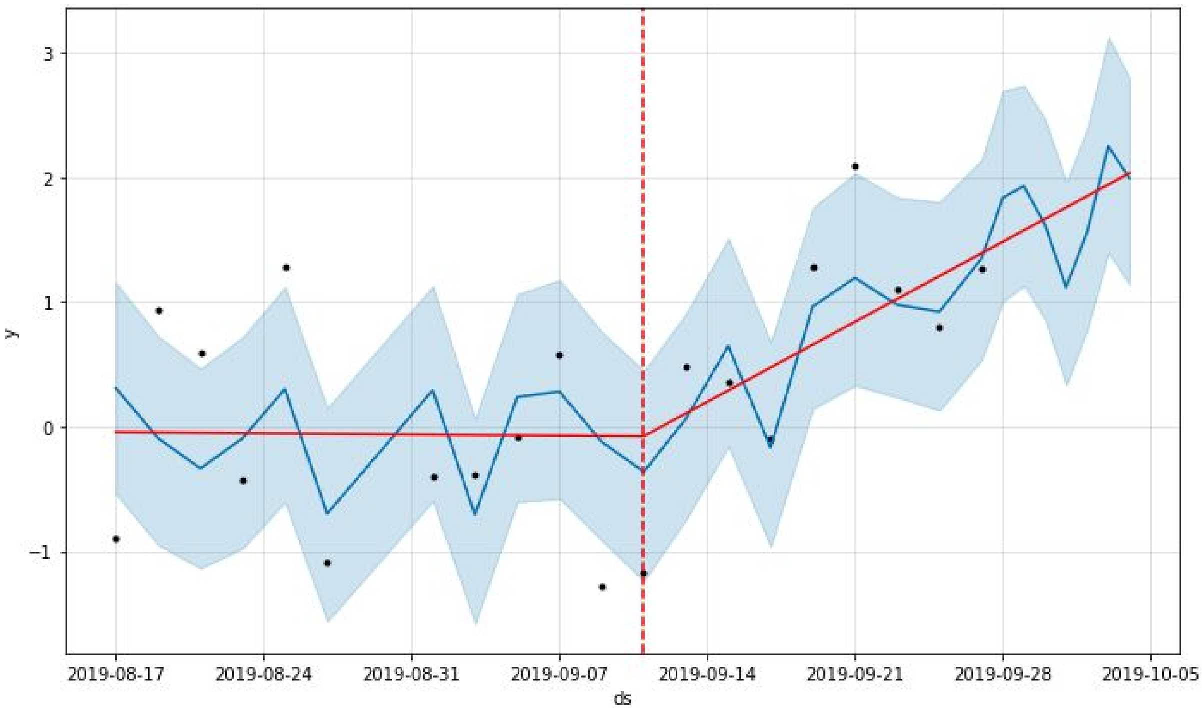
Figure 3: Using Prophet to Project Louisiana's Average Temperature to 2027



Using this regression model, we proceeded to make a model to project the annual mean loss proportion for Louisiana's rice industry. We proposed two models.

First was a Ridge linear regression model that takes the 16 weather parameters and maps them to the respective year's loss proportion. To encourage a nonlinear nature, which is critical to making a robust mathematical model, we fed the weather parameters into a second order polynomial feature generator. This generator takes in features, say [a, b, c], and manipulates them to return [1, a, b, c, a^2, b^2, c^2, ab, bc, ca]. This generator increases the number of features our model has to work with and can possibly give rise to some interesting interactions between various weather variables.

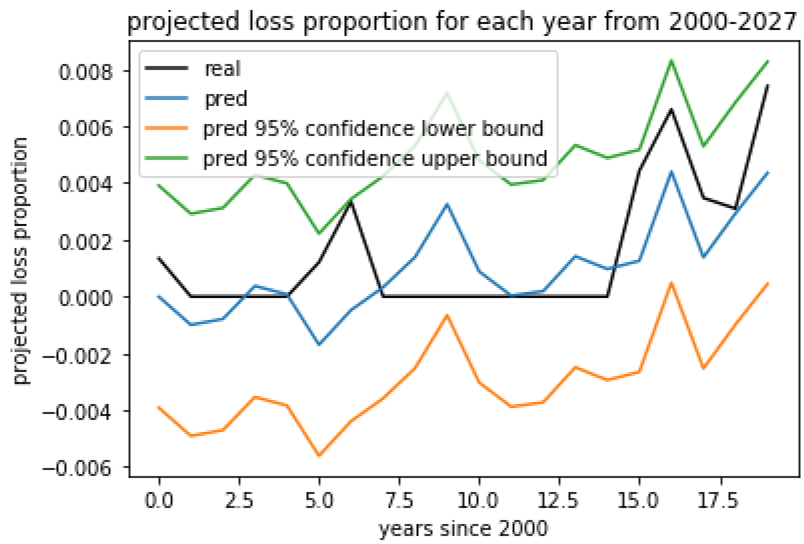
Figure 4: Date vs the Precipitation Index.



Tuning the flexibility coefficient gives the regression an inflection point

Unfortunately, when we fed our projected weather variables into our Ridge regression model, which used the results of the polynomial feature generator to run a 360 dimensional linear regression, we only got an r^2 score of 0.326. **Figure 5** shows the projection.

Figure 5: Projected Loss Proportions Using Ridge Regression



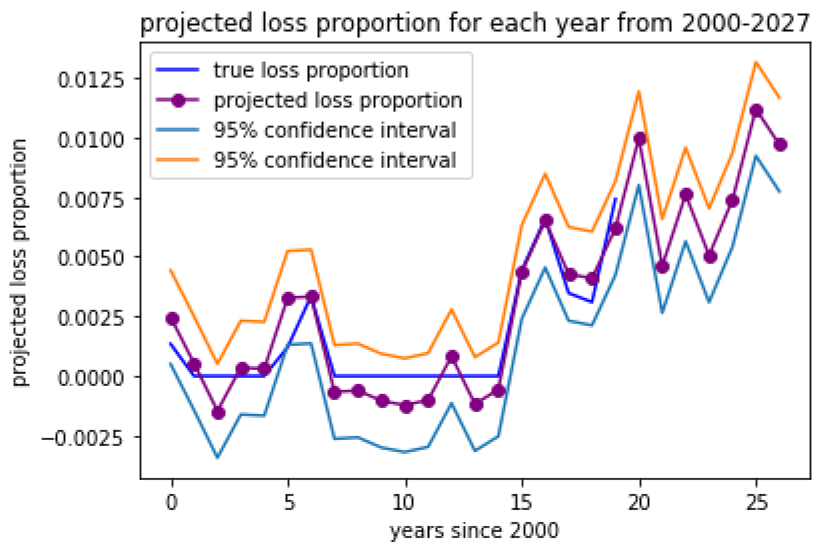
We used empirical testing to determine that ridge regressions gave better performance than lasso or elastic net, two other classical methods for optimizing a regression model. We also tested out deep neural network approaches but failed to get results that were better than the ridge regression.

After testing these models, we realized that our models were still black boxes. We could not interpret them so we decided to devise our own formula and optimize them with nonlinear least squares and a second order optimizer. After iteratively testing various nonlinear functions like sigmoid, 1/x, ln(x), we arrived at:

$$y = \sum_{i=1}^{16} log(|x_i|) * x_i * c_i .$$

Each x_i represents the weather parameters and the *c* coefficients are optimized with a soft L1 loss. Calculating *y* will give us the loss proportion. We achieved an r^2 score of 0.83 and the projections can be seen in **Figure 6**.

Figure 6: Loss Proportion Projection using Nonlinear Least Squares Optimization and Our Proposed Formula



In this paper, we proposed two robust models for both projecting weather data like the average temperature and various drought indices. We also provide a formula for predicting the loss proportion for a certain year given data describing its climate.

As such, the general trend is one of increasing loss proportions. Although there may be more arable acres of land that can be used to cultivate rice, each acre is more vulnerable to extreme weather events and the effects of drought on the aforementioned loss proportion.

In the short term, this will lead to more available land to grow rice on. Factoring in the many temperature demands of the rice through its maturation process, the ideal temperatures for rice cultivation is between 25℃ and 38℃. We the actuaries are not worried about temperatures going below 25℃, as our model shows that average temperature has generally been increasing and the rice insurance claims due to frosts or cold weather has been as well. However, temperatures over 38℃ are concerning as hotter conditions would lead to sterile pollen.[5] Another metric to consider, which we did not have the data for, is average nighttime temperatures. Ziska and Manalo of the University of Arkansas corroborate that higher nighttime temperature leads to grain yield and the yield itself consists of smaller and chalkier rice that would be seen as subpar in quality compared to similar strains produced elsewhere.[6]

In the long term, however, net output will decrease along with an increase in the loss proportion. After all, the USDA's temperature range for rice cultivation does not exceed the aforementioned 38℃. Our average temperature regression shows that by 2026, the average temperature in Louisiana will exceed 38℃. If the average acre of land is suboptimal for rice growth, the loss proportion should be high as well, and in 2026, our model predicts that it will be between 0.01002 - 0.01455. This is a 65% increase in the loss proportion from that of 2019 (0.00741).

---

[5] (2013, February). Retrieved from https://www.usda.gov/oce/climate_change/effects_2012/CC and Agriculture Report (02-04-2013)b.pdf
[6] (Cooper, Siebenmorgen, & Counce, 2008)

With our findings, one recommendation for farmers would be to integrate more advanced technology into their farming practices to mitigate their losses. Cultivating modified rice crops would significantly increase yields. According to a 2018 report by the United States Department of Agriculture, Southern rice producers have been increasingly planting new rice seed varieties since 2000, including hybrid and non-genetically modified herbicide-tolerant seed.[7] Furthermore, the development of genetically modified drought-resistant strains of rice make it increasingly possible to cultivate the crop in locations with unreliable water supply. The USDA also notes that one of the most important increases in precision farming technology these past two decades has been the growth of guidance systems. Used to steer tractors and other self-propelled machines, these systems are more accurately able to reduce any coverage gaps and overlaps. By 2013, over 50 percent of rice producing acreage employed automatic guidance systems, allowing farmers to water their fields much more efficiently. Similarly, drones or other sensors deployed in the air are able to capture data and predict how crops should be watered. These precise farming technologies and guidance systems, in addition to modified crops, have had significant impacts on US rice production; between 2000 to 2013, productivity was estimated to have increased by 29%. Ultimately, adopting these technological advances in agricultural production will allow farmers in Louisiana and across the US to increase their crop yields and compensate for forecasted losses due to climate change and poor water access.

We also recommend that insurance companies implement some sort of way to adjust their policy calculations based on the technology that a farmer might possess. In the near future especially, technology will only get better at mitigating losses in agriculture, which is a major

[7] McGinnis, M. (2018, August 3). Over Half of U.S. Farms Lose Money, USDA Study Shows. Retrieved from https://www.agriculture.com/over-half-of-us-farms-lose-money-usda-study-shows

input to how insurance policies are determined: the lower the probability of loss, generally speaking, the lower the insurance premiums paid by the insured to the insurer.

However, the affordability of more sophisticated technology for agricultural purposes is still an obstacle for many small farmers. In fact, many are unable to support their farms even without these extra expenses due to severe weather and falling prices. According to a 2019 report by TIME magazine, “farm bankruptcies were up 12 percent in the Midwest from July of 2018 to June of 2019; they’re up 50 percent in the Northwest. Tens of thousands have simply stopped farming, knowing that reorganization through bankruptcy won’t save them. The nation lost more than 100,000 farms between 2011 and 2018; 12,000 of those between 2017 and 2018 alone.” A statement by the USDA’s Economic Research Service revealed that more than half of the roughly 2 million US farm households lose money from their farming operations each year. With these dire conditions, the futures of small farmers amid weather events and crop losses that are only projected to get worse remain uncertain.

Subsidies are another means of offering an incentive for farmers to adopt cutting edge innovation to minimize losses. In the same manner that a rebate is offered for installing solar panels, the federal government needs to provide a tax cut or rebate for adopting well tested and effective technology that mitigates the effects of climate change on the loss proportion and yield of many crops, including rice.

Naturally, small farmers don’t have the funds to do that, as the rebate would only cover a small percentage of the total cost of the technology implementation, much like solar panel rebates.[8] One way to bypass this issue and to enable them to stay afloat is to provide them with

---

[8] New York Solar Rebates And Incentives. (n.d.). Retrieved from https://www.cleanenergyauthority.com/solar-rebates-and-incentives/new-york

subsidies for implementing a crop rotation, where, hypothetically, they grow rice one year and legumes or some crop good for the soil the next. As a result, the small farmers can do a crop rotation to put nutrients back into the soil for resource intensive crops such as rice. Without the subsidy, they couldn't rotate as farms are economies of size and the bigger it is, the smaller the next size increase will cost relative to the past one.[9] By cycling amongst crops, especially in an area such as the Mississippi Delta, where the rate of rice monoculture (no cycling of crops) is 40%, the rice will have more nutrients in the soil, have a higher productivity per acre, which, in theory, can compensate with the natural loss of rice acreage productivity due to increasing temperature.

[9] Duffy M. (2009). Economies of Size in Production Agriculture. *Journal of hunger & environmental nutrition*, *4*(3-4), 375–392. https://doi.org/10.1080/19320240903321292