Big Data = Big Future + Big Challenges

Larry A. Larson Speaker Series Event
held as a plenary session
at
the FMA Annual Conference
Reno, Nevada
on
Sept. 5, 2018

Keynote Speaker: Daniel Cotter
Panelists:

Panelist Talking Points

David Ford, HDR David Ford Consulting

- We floodplain managers have (and need to acknowledge and work to overcome) cognitive biases. These biases affect our ability to judge the value of big data. We are mostly scientists and engineers, fascinated with numbers and analyses. That fascination leads us to desire success with large-scale hazard and consequence analyses. The desire draws us to details that confirm existing beliefs and patterns that demonstrate success even when we evidence is sparse. We may get the right answer for all the wrong reasons, an outcome we must seek to avoid.

- DATA is not INFORMATION. We make decisions based on INFORMATION. We apply KNOWLEDGE to DATA to create INFORMATION. In my opinion, our greatest concern as floodplain managers should be the KNOWLEDGE application. What models and methods do we use to apply KNOWLEDGE? And who does that? Are they qualified? Do they understand the relevant processes? Or are they getting the right answer for the wrong reasons?

- Applying KNOWLEDGE at a fine temporal and spatial scale with poor quality DATA does not produce better INFORMATION for DECISION MAKING. For example, the 2015 NRC study "Tying flood insurance to flood risk for low-lying structures in the floodplain" concludes "Modern technologies, including analysis tools and improved data collection and management capabilities, enable the development and use of comprehensive risk methods, which could improve NFIP estimates of flood loss." The report goes on to suggest we could "…determine the flood hazard for individual structures…" Recent analyses in California demonstrate this fine resolution analysis is feasible: Economic flood risk for approximately 400,000 structures was estimated and used to inform plan formulation, and vendors provide similar resolution analyses. But a 2009 NRC study ("Mapping the zone") points out "Topographic data are the most important factor in determining water surface elevations, base flood elevations, and the extent of flooding,

and, thus, the accuracy of flood maps in riverine areas." And we all are aware of the limitations of those data—particularly the National Elevation Dataset, etc. So if flood risk is analyzed by structure, but the analysis is based on elevation DATA that are inaccurate, the resulting INFORMATION will not be inaccurate, and the decisions made may be wrong—wrong enough to put people at risk or to incur unnecessary costs for risk management.

- Applying poor or outdated KNOWLEDGE to massive DATA sets does not yield massive INFORMATION. For example, a paper published in 2015 in the Journal of Hydrology addressed the question "How much can we gain with increasing model complexity with the same model concepts?" The finding was "…there is no superiority of complex grid-models over simple grid models in reproducing internal variables in this study." Another paper in the same journal addressed the question "Does model performance improve with complexity?" The authors found …added complexity does not necessarily lead to improved performance of hydrological models.

Samantha Medlock, Willis Towers Watson

- As Big Data increasingly play a role in risk management, and in the quantification of risk, questions arise regarding how insurers and reinsurers are making use of Big Data.

- For insurance purposes, big data refers to unstructured and structured data being used to influence underwriting, rating, pricing, forms, marketing and claims handling. Structured data refers to data in tables and defined fields and datasets. Unstructured data, comprising most data, refers to things such as social media postings, digital images and video, and information transmissions such as text and instant messages. Predictive analytics allows insurers to use big data to forecast future events. The process uses a number of techniques—including data mining, statistical modeling and machine learning—in forecasts.

- The 2016 Predictive Modeling Survey conducted by Willis Towers Watson indicates that Big Data, from vehicle telematics and the IoT, are opening up many new potential avenues for improvement, as personal auto carriers expect to get much more driving data from connected cars (100%), apps (75%) and telecoms (63%) in the next five years. P&C insurers expect big data use in many key business areas (such as claim management, understanding customer needs and product development) will more than double in the next two years. Expected sources of these data are both internal and external, including telematics, web clickstreams, customer-agent interactions, smart home data and social media, accompanied by a shift toward greater use of machine-learning techniques.

- While the use of big data can aid insurers' underwriting, rating, marketing, and claim settlement practices, the challenge for industry and regulators is to examine whether it is beneficial or harmful to consumers. Concerns have arisen relating the lack of transparency and potential for bias in the algorithms used to synthesize big data, as well as the collection of information that may be sensitive to consumers' privacy or potentially discriminatory.

- Similar to the use of credit scores in pricing life insurance or "redlining" areas based on race, it will be important to distinguish between factors where there may be an apparent correlation but no actuarial bases for interpreting data as they may pertain to immutable characteristics, such as race or gender, or to income, credit, or other social factors. This is also important in the context of flood risk because it is already well documented that lower-income property owners and communities suffer disproportionately higher flood risk across the U.S. – and indeed globally.

- There have been important applications in the use of Big Data to help enhance understanding of flood risk and to map flood extents, for example using Twitter for real-time information about where users are experiencing and document flood conditions and the use of Google platforms for mapping floor elevations.

- In addition to flood risk mapping, among the most promising areas for future development are 1) mapping and understanding erosion risk; 2) mapping and quantifying the value of natural resources and the benefits they can provide in attenuating surge and flood impacts and addressing stormwater problems; and 3) in improving methodologies and data inputs to benefits-costs analyses for flood hazard mitigation to ensure the full range of social and economic benefits are included.

- It will be important to develop broad-based principles and guidance for use of big data to prevent the misuse of data and to help ensure transparency and the appropriate use of Big Data in flood risk management and insurance practices.