

**Testimony of  
Bert Little, Ph.D.  
Associate Vice President for Academic Research  
Professor of Computer Science and Mathematics  
Executive Director, Center for Agribusiness Excellence  
Tarleton State University  
Texas A&M University System  
Before the Subcommittee on General Farm Commodities and Risk Management  
Thursday, June 7, 2007  
10:00 AM**

Chairman Etheridge, Ranking Member Moran, and Members of the Subcommittee, thank you for the opportunity to appear this morning before the Subcommittee to discuss efforts to eliminate fraud, waste, and abuse in the Federal Crop Insurance program and, in particular, the data warehousing and data mining efforts currently undertaken by the Center for Agribusiness Excellence (CAE) with the Risk Management Agency (RMA).

**I. Center for Agricultural Excellence (CAE)**

I am Bert Little, Associate Vice President for Academic Research and Professor of Computer Science and Mathematics at Tarleton State University, which has been a member of the Texas A&M University System since 1917. In this role, I also direct the CAE, which was founded at Tarleton specifically to address the section of the Agriculture Risk Protection Act of 2000 (ARPA 2000) directing the Secretary of Agriculture to use data mining and data warehousing to improve integrity and compliance in Federal Crop Insurance. Personally, my own roots in agriculture run deep. My family obtained its first land grant in 1790 in southeastern North Carolina, and I worked on that same piece of land raising tobacco, corn, and soybeans until my late teens.

Recent press reports as well as testimony before this and other Committees of Congress have raised questions about the integrity and cost-efficiency of the Federal crop insurance program, and I appreciate the action of this Subcommittee in devoting this morning's hearing to this topic of great concern both to farm producers and to taxpayers in general. I will use my testimony to give the Subcommittee a fresh update on the program integrity activities conducted by USDA's Risk Management Agency (RMA) through its data mining and data warehousing initiatives housed at CAE. At the outset, I would emphasize that the data analyzed at CAE involves insurance policies and potential fraud and abuse of those policies. We do not analyze the financial data involving crop insurance companies who deliver the program to producers under the Standard Reinsurance Agreement with FCIC. As a result, my testimony will not address those issues.

## **II. DATA MINING AND WAREHOUSING OVERVIEW**

We are pleased with the success CAE has had in applying data mining techniques to the crop insurance program. USDA's Risk Management Agency, in its annual Program Compliance and Integrity reports to Congress, has conservatively estimated that, over a period of six years, we have saved American taxpayers nearly half a billion dollars by highlighting potential fraud and abuse in the program and, as a result, helping RMA to avoid making improper payments. These savings are detailed in Appendix I. RMA and its staff are to be complimented for their effective and aggressive use of these powerful new compliance tools.

In the course of our analytical work, we have found that the farmers who participate in the Federal crop insurance program by and large are honest people who follow the rules. Our Spotcheck program, described in more detail below – designed to identify suspicious patterns indicating possible program abuse – has consistently found fewer than one percent of producers falling in this category. This is a strong indicator of program integrity and rates much better than comparable lines of insurance in the property and casualty field.

Crop Insurance is a data intense program with complex rules. Data mining works well in such an environment. The savings we have accomplished were achieved through a variety of coordinated activities aimed at exposing and preventing abuse. As a starting point, CAE has built a data warehouse comprised of all RMA policy information from 1990 to the present, as updated by RMA every two weeks. Data on weather, soils, and other agronomically relevant factors are integrated into the CAE data warehouse to complement policy data for analysis. All data maintained with the CAE data warehouse are subject to the same USDA privacy and security protections that apply to data maintained by USDA itself. The result today is a database containing more than two terabytes (terabyte = 1 trillion words (bites) of information, and we have linked this data across time to allow multiyear comparisons, a key analytical approach previously unattainable. With it, CAE produces more than 100 data mining research products each year in coordination with USDA.

## **III . ONE EXAMPLE: THE SPOTCHECK LIST**

One example of how we use this foundation to identify and prevent abuse is the system that CAE has developed, along with RMA staff, to use its database each year to identify multi-year patterns that signal suspicious or anomalous crop insurance claims. The result is what we call the Spotcheck List, an actual list of producers who will then become subject to increased compliance oversight. Over the years, we have refined this process to the following five steps:

- (1) Based on such starting points as anecdotes from the field or experience of investigators, producers, agents, or adjusters about schemes to exploit the program, we design data mining algorithms to identify schemes that farmers might potentially use to obtain improper crop insurance indemnities;

- (2) These schemes are analyzed to determine whether they occur in the national data, where, and to what extent. RMA and CAE analysts review the data mining analyses to determine whether or not the scheme is structured and results in personal benefit;
- (3) Schemes and specific producers are identified and placed on the Spotcheck List. The List is reviewed by RMA Compliance staffers, who may add additional persons of interest to the List;
- (4) The Spotcheck List is passed to USDA Farm Service Agency (FSA) State Executive Directors, who ask local county FSA offices to conduct inspections during the growing season on the identified fields;
- (5) FSA sends a letter to each producer on the Spotcheck List notifying them that an inspection will be performed on his or her crop. Additional pre-harvest visit(s) may be made.

Most producers on the Spotcheck List react to the FSA letter in step 5 by refraining from any contemplated abusive activities. The result is a visible, measurable reduction in indemnities paid. Simply put, growers change their behavior as a result of knowing that they are being scrutinized. Before they were on the Spotcheck List, this subgroup of producers had loss ratios that were several fold higher than their neighbors in their own counties. But after being informed they were on the list, their loss ratios fell to the county averages. Importantly, this effect of reduced indemnity lasts several years among more than two thirds of those on the Spotcheck List. In sum, over six years (2001 through 2006), the Spotcheck List initiative alone has produced measurable reductions in unneeded indemnities of approximately \$479 million .

As noted, the CAE Spotcheck List is only one of more than 100 research products produced annually by CAE at the request of RMA aimed at improving program integrity and contributing materially to cost savings. Other federal offices that have requested and received assistance from CAE in the form of data mining analyses have included the USDA Office of Inspector General (OIG), the Government Accountability Office (GAO), and various Federal prosecutors as well as investigators from the Federal Bureau of Investigation (FBI). When requested, CAE personnel have served as expert witnesses for Federal prosecutors in crop insurance fraud litigation.

#### **IV. DATA SHARING; ONGOING RESEARCH AND FUTURE DEVELOPMENT**

Our current analytical products can and should be more fully utilized, and we believe the next logical extension would be to better include in the process the reinsured companies who deliver crop insurance to producers across the country. At a meeting of the National Crop Insurance Services (NCIS) organization last month, RMA Administrator Eldon Gould announced that these insurance companies will now begin having access to the CAE Dashboard – our basic platform for accessing data at the county level -- and they will be able to submit work orders directly to CAE for research on specific problems. At the same meeting, I was given a green light to announce that the insurance companies would be provided a secure portal through which to report their growing season inspections on the RMA SharePoint system, the system they currently use to transmit policy information to RMA's Kansas City Data Center.

One analytical tool available on the CAE Dashboard that offers a particularly powerful resource is our searchable, stored archive of NEXRAD weather loops – essentially the same Doppler Radar images we see on our local television weather reports. To our knowledge, CAE maintains the only such active system of NEXRAD data maintained over a period of years. In one example, for instance, two farmers filed claims on hail damage that were denied because NOAA could not verify that a hail storm had occurred on the day in question. But by using our NEXRAD system, we were able to identify a very isolated, very heavy storm that produced the damage. As a result, the farmers' claims were verified, and they could be paid the indemnity they deserved.

Most recently, CAE, in collaboration with the Stennis NASA Space Center Applied Sciences, has completed much of the process of integrating into the data mining process satellite data that measures the intensity of the green light reflected by the chlorophyll molecules in plants – a measure of biomass present. CAE has invested its own non-Federal resources to build a 42 Terabyte data system to store or hold our satellite data for January 1, 2000 to the present. Our preliminary results are exciting, indicating a better than 90 percent ability to evaluate crop production via satellite using this system, and we are currently working to augment it with data from the Indian AWiS satellite.

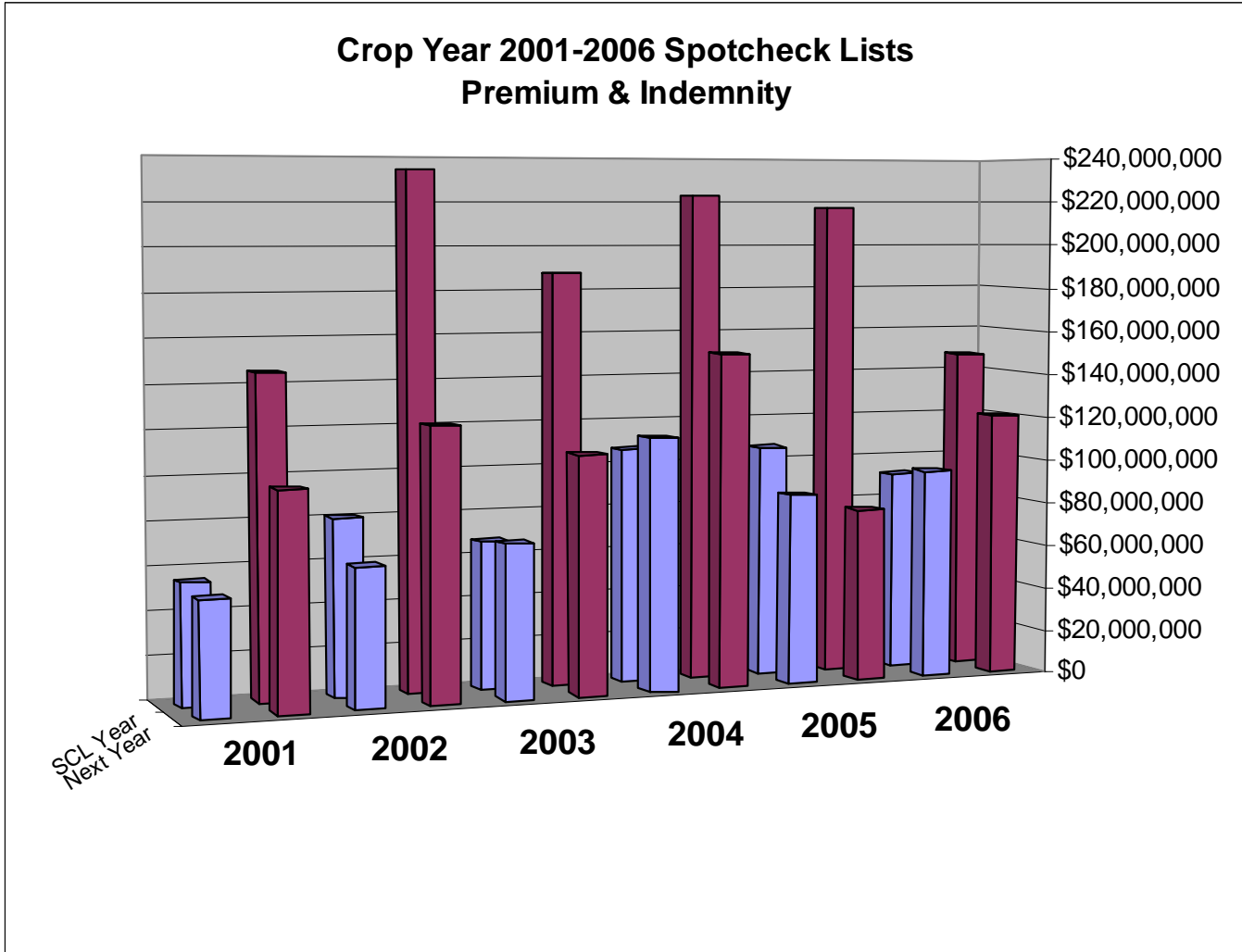
In the future, CAE hopes to incorporate in our system the Common Land Unit (CLU) data held by USDA's Farm Service Agency. With CLU data, we will be able to assess biomass health (indicated by reflectance of chlorophyll green) at the field level using satellite data and quantify its direct relationship with crop production. We see many such opportunities to improve our analysis with the inclusion of farm data reported to FSA, and we have been requesting FSA to provide it to us for this purpose.

## **V. SUMMARY**

Data mining as mandated under ARPA 2000 has been a striking success for Congress and USDA. For an investment of \$26.1 million, it has conservatively produced program savings of over \$479 million since December 2000 with the Spotcheck List alone. For the longer term, Congress may wish to consider continuing this program by providing a multi-year funding authority in the 2007 Farm Bill, similar to the multiyear approach used so effectively to fund the program in the original 2000 APRA legislation.

Thank you again for giving us this opportunity to summarize CAE's record of providing cost savings to the Federal Crop Insurance program under the ARPA 2000 data mining program. Congress and USDA deserve a great deal of credit for taking the bull by the horns and implementing this program in an effective way to the benefit of both farmers and taxpayers. Great strides have been made to improve the policing of the Federal crop insurance program since the adoption of ARPA in 2000, and we have been honored to be part of the process. Thank you for your consideration, and I would be happy to answer any questions you might have.

## Appendix I



**FIGURE: Indemnity Decreases for 2001-2006: \$479 Million.**  
 Taller bars (maroon, back row) reflect payments before producers were on the Spotcheck List, and the shorter bars (maroon, front row) are after they were on the Spotcheck List.