The Designing of the New Regression Models of Crop Productivity Year-to-Year Anomalies Based on the AVHRR Satellite Vegetation Monitoring Information

Menzhulin G.V, Peterson G.N., Pavlovsky A.A.

Research Centre for Interdisciplinary Environmental Cooperation of Russian Academy of Sciences

Kogan F.N.

US NOAA, NESDIS, Maryland

Symposium on Climate Change and Variability-Agro Meteorological Monitoring and Coping Strategies for Agriculture

Oscarborg, Norway
June 3-6, 2008

Principal Purposes of Assessment

- To assess the perspective of satellite vegetation indices data using as the numerical predictors of crop productivity regression models.

- To develop the regression models *"Crops productivity anomalies – Vegetation Indices"* for representative agricultural areas having the high territorial resolution of the crop productivity information.

- Basing on the results of numerical experimentation to determine the features of such kind of crop productivity models.

Information on Vegetation Indices Temporal Series and Year-to-Year Crops Productivity Predictants

AVHRR, 16x16 км Resolution: NOAA, NESDIS, 1982-2005:

Temperature Condition Index

$$TCI = 100(BT_{max} - BT)/(BT_{max} - BT_{min})$$

BT = Brightness Temperature

Vegetation Condition Index

$$VCI = 100(NDVI-NDVI_{min})/(NDVI_{max}-NDVI_{min})$$

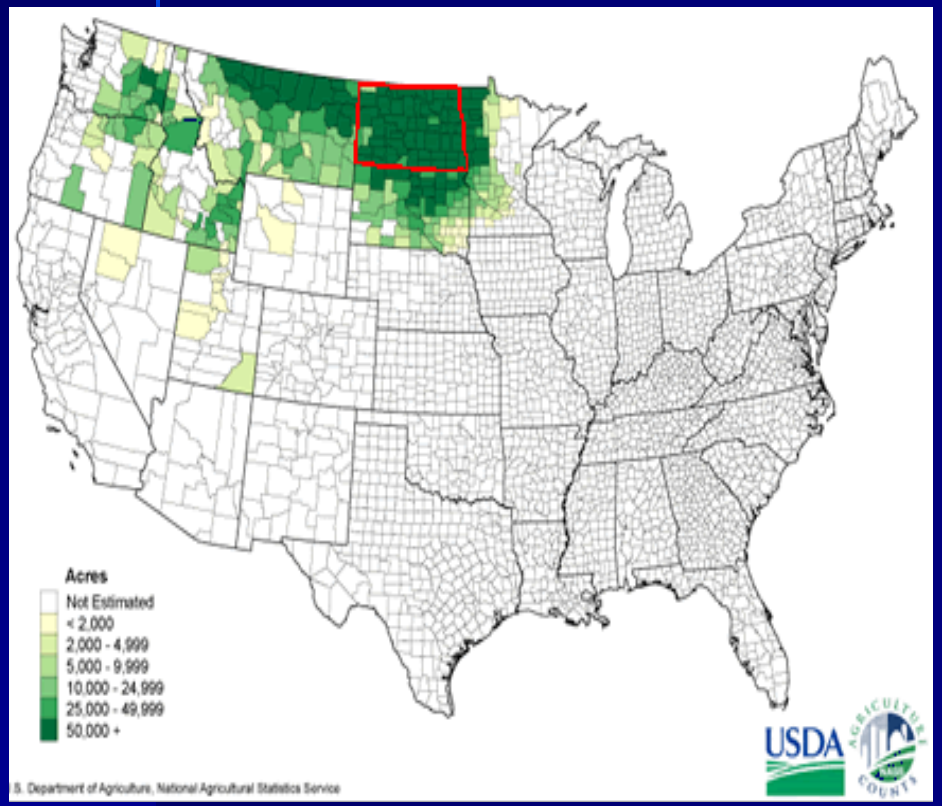
NDVI=(VIS-NIR)/(VIS+NIR)

Crops Productions and Areas:

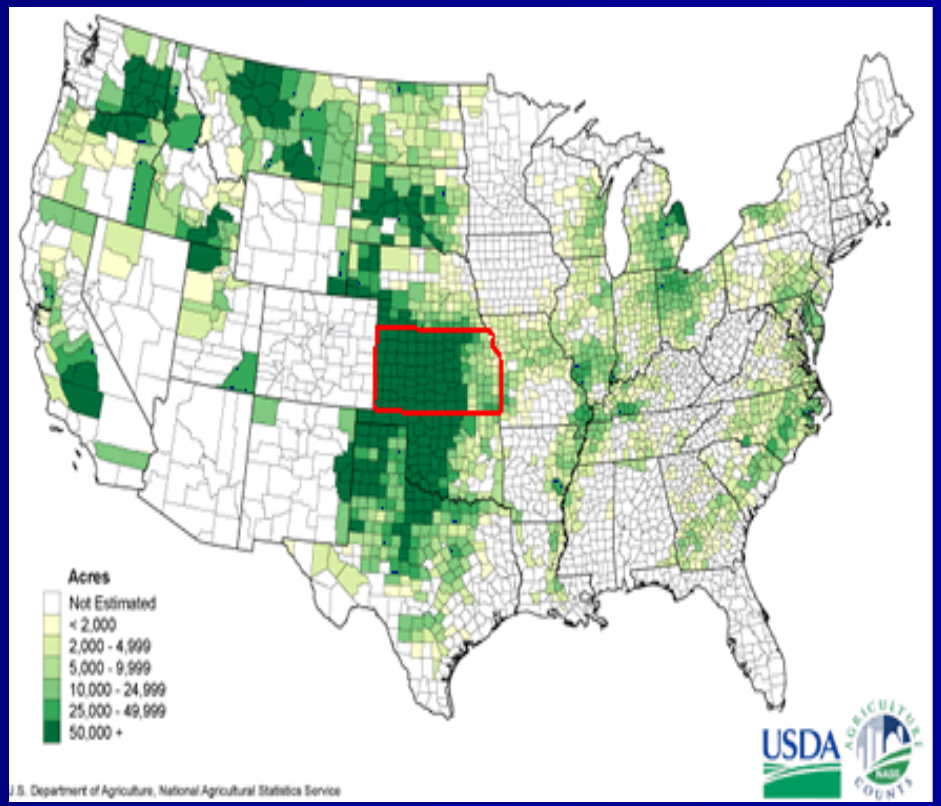
USDA, 1982-2006, wheat production and areas in US counties (more 3000)

Areas of Spring and Winter Wheat Production over the US Territory

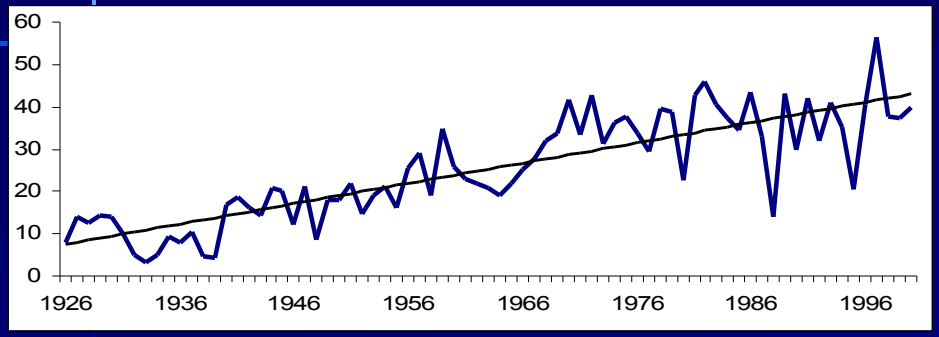
США, 2005



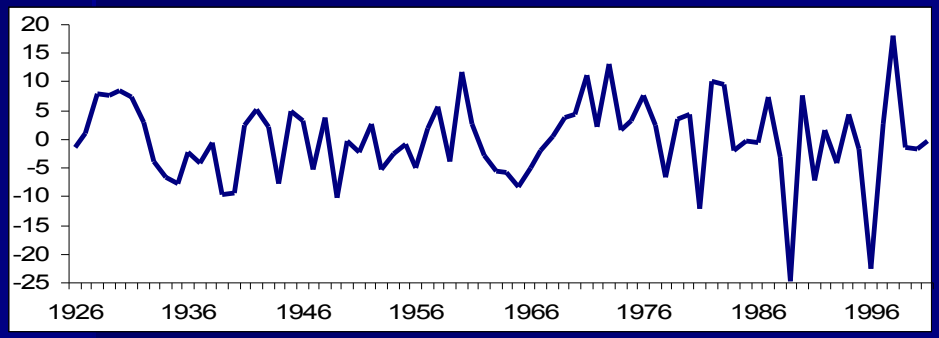
Посевы озимой пшеницы, США, 2005



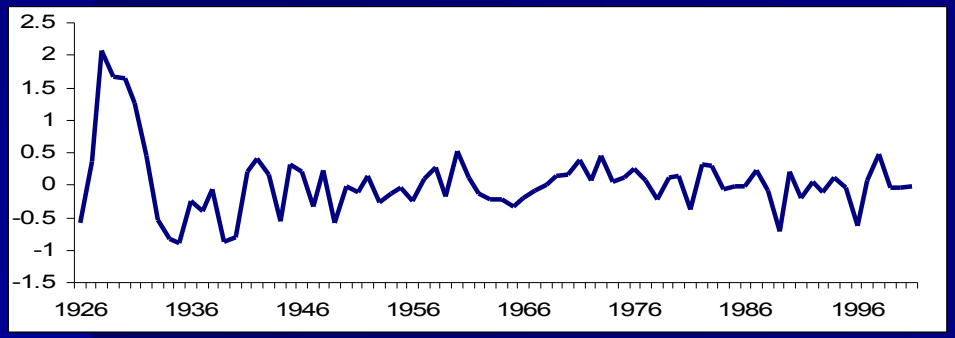
Temporal Dynamics of Crop Productivity Indicators



Real dynamics of productivity (without trend extraction)



Dynamics of absolute anomalies



Dynamics of normalized anomalies

Examples of Linear Trends in Vegetation Indices Series

Counties Kansas	VCI												TCI											
	Months												Months											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Harvey	2	1	1	0	1	4	4	3	1	1	0	1	-2	-1	-1	-1	-1	0	1	0	-2	-2	-1	-1
Haskell	1	1	0	0	1	4	6	6	3	1	0	1	-1	0	-1	-1	-2	0	1	1	-1	-1	-1	0
Hodgeman	2	2	1	1	1	1	1	2	2	3	2	2	-1	0	0	-1	-1	-1	0	0	0	-1	-1	-1
Jackson	3	4	3	2	3	3	3	2	2	3	2	2	-2	-3	-2	-1	-1	0	1	-1	-2	-2	-2	-1
Jefferson	3	3	2	2	2	4	4	2	2	3	2	2	-2	-2	-1	-1	-1	0	1	0	-2	-2	-2	-1
Jewell	2	1	0	0	1	2	2	2	1	1	1	2	-1	0	-1	-2	-2	-1	1	0	-1	-2	-2	-1
Johnson	3	3	3	2	2	2	2	1	2	3	2	3	-2	-2	-2	-2	-1	-2	-1	-1	-2	-2	-2	-2
Kearny	2	2	1	1	1	3	4	4	4	3	2	2	-1	0	-1	-1	-1	0	2	1	0	-1	-1	0
Kingman	2	1	0	0	1	3	3	3	2	1	1	2	-2	-1	0	-1	-1	0	1	0	-2	-1	-1	-1
Kiowa	1	1	-1	-1	1	3	2	2	1	1	1	1	-1	0	0	-1	-1	0	1	0	0	-1	-1	0
Labette	2	2	2	2	2	5	5	3	2	2	2	2	-2	-1	-1	0	0	0	1	-1	-3	-2	-1	-1
Lane	2	2	0	0	1	2	2	3	3	2	1	2	-1	0	-1	-1	-1	-1	0	0	-1	-1	-1	-1
Leavenworth	3	2	2	1	2	3	3	2	3	3	2	3	-2	-2	-1	-1	-1	-1	0	-1	-2	-2	-2	-1
Lincoln	2	2	0	0	1	3	2	2	2	3	2	2	-1	-1	-1	-2	-1	0	1	0	-1	-1	-2	-1
Linn	3	3	3	2	3	4	4	2	2	3	2	3	-2	-2	-2	-1	0	0	1	0	-1	-2	-2	-1
Logan	3	3	1	1	1	1	2	3	3	3	2	2	-1	0	0	-1	-1	-1	0	0	0	-1	-1	-1

0-No trend, 1-Very slight, 2-Slight, 3-Moderate, 4-Significant, 5-Very significant. **Red-positive, Blue-negative**

A. Types of Statistical Model Used

1. Semi-Empirical Models

Stages of designing

- The first (leading) predictor is chosen as maximally correlated one with predictant (dependent value)

***Other five predictors are chosen according the following*:**

- Second predictor - next maximally correlated with predictant but which correlated with the first (leading) predictor not higher than 0.9.
- Third predictor - the next maximally correlated with predictant but which correlated with the second and leading (first) predictor not higher than 0.9.
- After five such steps the five predictors additional to the leading one shell be chosen.

Finally, using the predictors chosen in the procedure of back run analysis, permitting to delete consequently statistically less significant predictors, the final model provided the maximal value of adjusted coefficient of determination will be found

2. Models Based on the Statistical Significant Criteria

Stages of designing

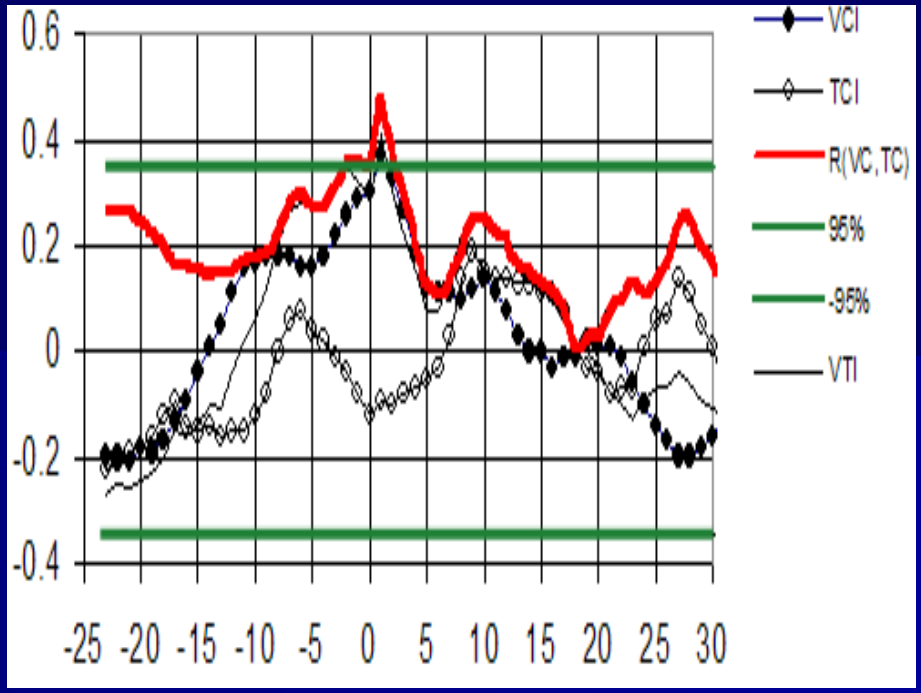
- The first (leading) predictor is chosen as maximally correlated one with predictant (dependent value)

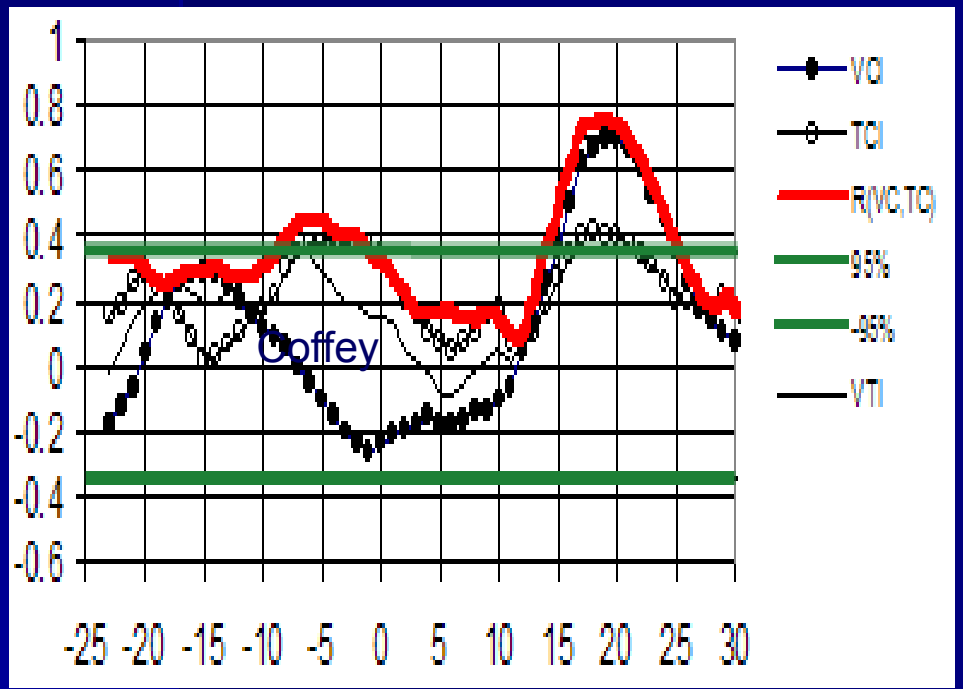
Other five predictors are chosen according the following:

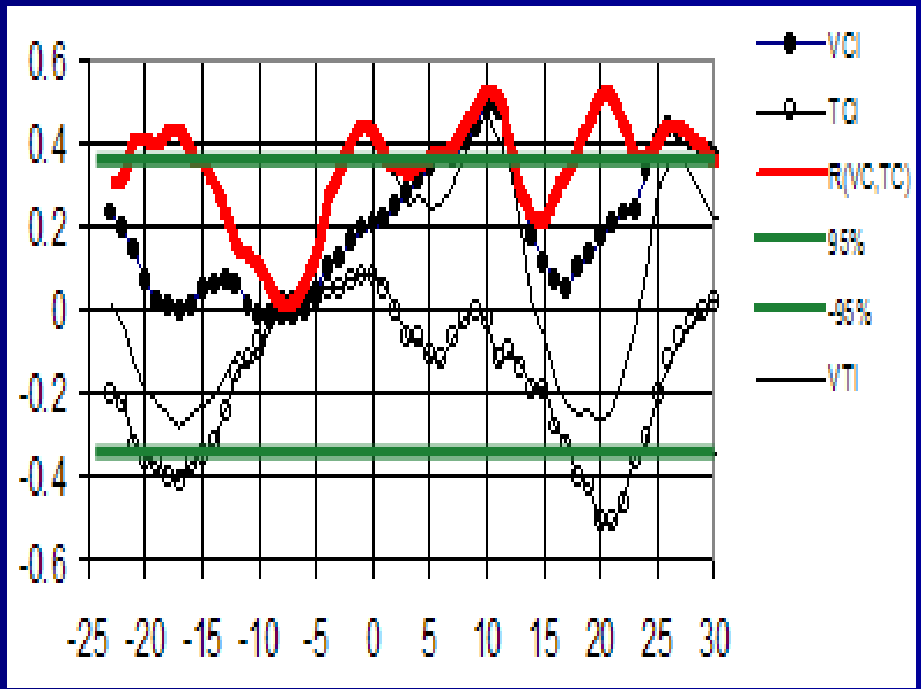
- Develop all (for growth period) two-factors models (leading predictor plus other one).
- Chosen the six models having the highest ratios of their coefficient of double correlation to the correlation coefficient of leading predictor.

Finally, those five predictors which provide the best (with the leading predictor) six double-predictor models used when developing of six-predictor model. The procedure of back run results. permitting to delete consequently statistically less significant predictors, the final model provided the maximum of adjusted coefficient of determination will be found

Semi-Empirical Technique of Vegetation Indices Selection when Regression Modelling







B. Types of Statistical Model Used

3. Exhaustive Models

Stages of designing:

Analysis of regression models developed on the principal of direct exhaustion from the complete set of predictors:
Amount of predictors per one point – 52 VCI & 52 TCI. Total - 105

- Two-predictors models:

Amount of analyzed models = (104·104) / (1·2); **About five thousands**

- Three-predictors models:

Amount of analyzed models = (104·104·104) / (1·2·3) = 1000000/3; **About three hundred thousands**

- Four-predictors models:

Amount of analyzed models = (104·104·104·104) / (1·2·3·4) = 100000000/24; **About five million**

- Five-predictor models:

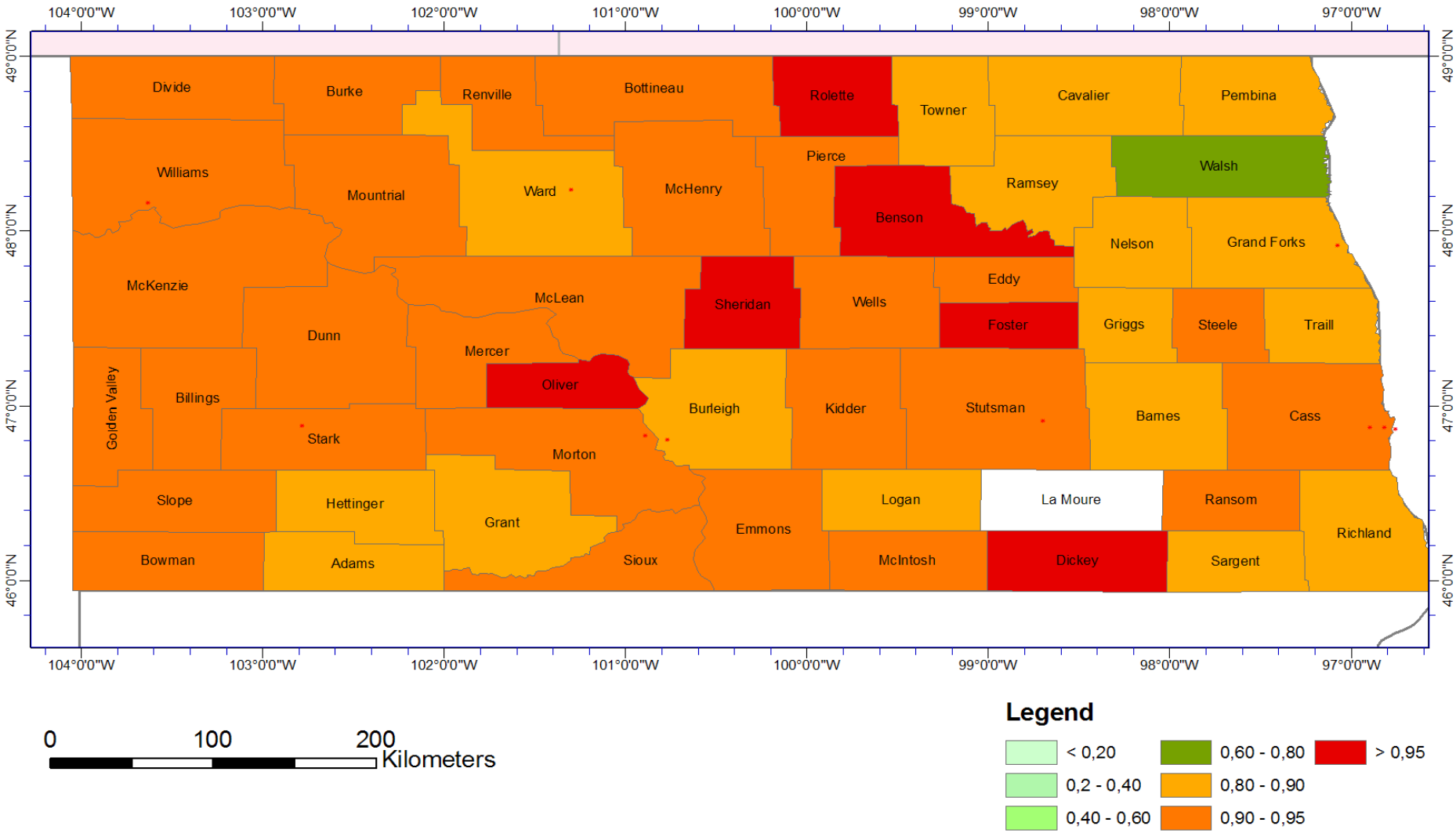
Amount of analyzed models = (104·104·104·104·104) / (1·2·3·4·5) = 10000000000/120; **About hundred million**

- Six-predictor models:

Amount of analyzed models = (104·104·104·104·104·104) / (1·2·3·4·5·6) = 1000000000000/720; **About thousand million**

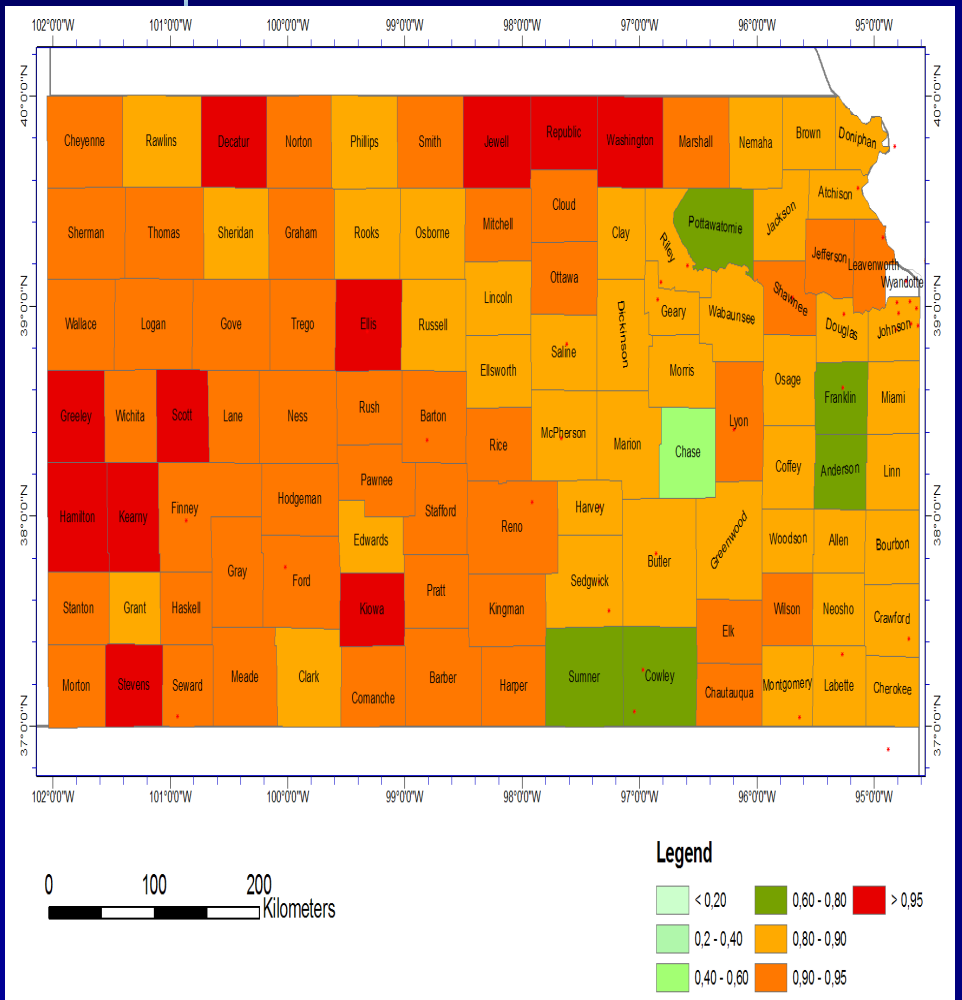
Coefficients of Determination of the Exhausted Five-factor Regression Models (quadratic trends extracted)

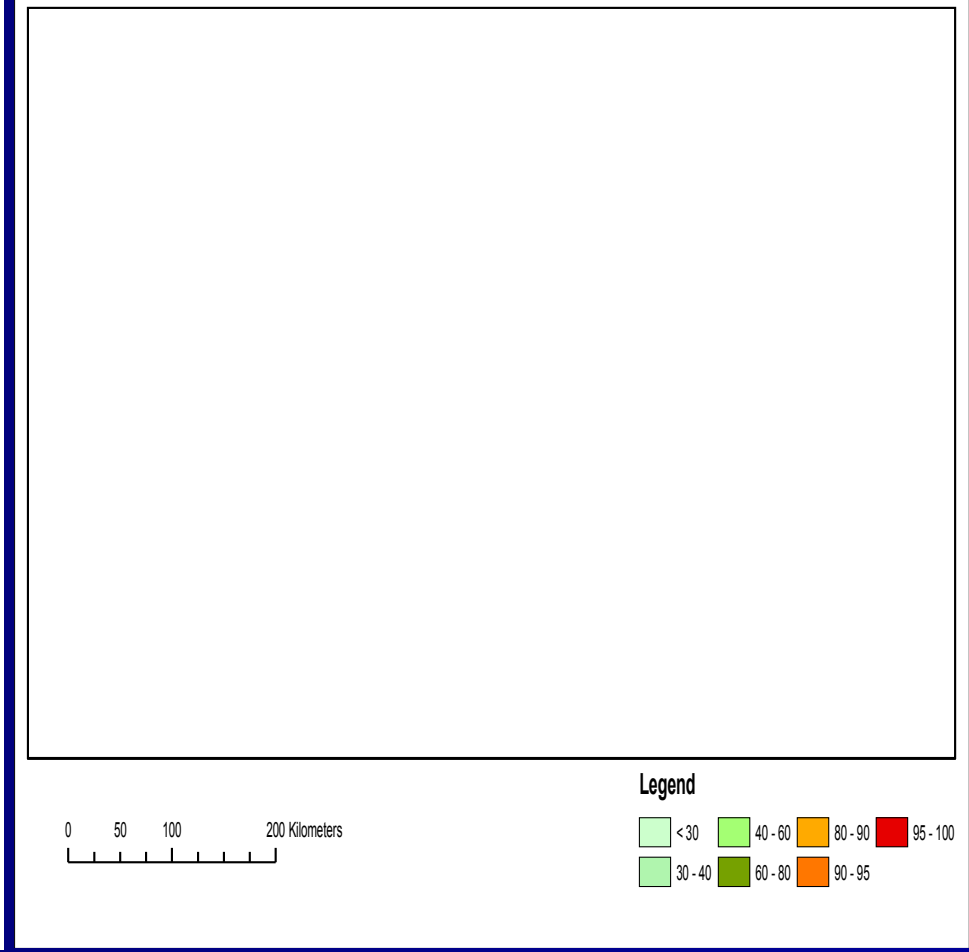
Spring Wheat, Counties of North Dakota



Coefficients of Determination of the Exhausted Five-Factors Models (quadratic trends extracted) (Left) and the same Coefficients for Models Based on the Monthly Surface Meteorological Data (Right)

Winter Wheat, Counties of Kansas





Two Examples of the Regression Models Parameters Calculated by Empirical (CR,PH) and Exhausting (EX) Techniques

Cheyenne

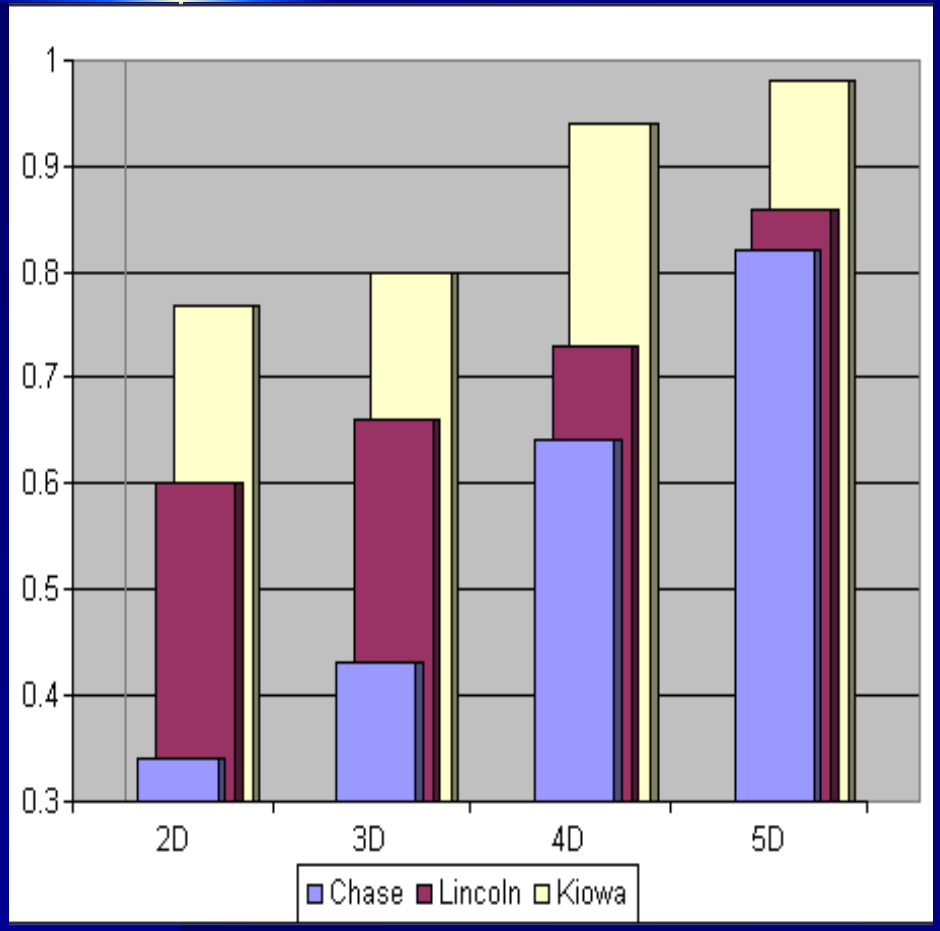
Model	Predictors and their t-criteria										D	Da	σ
	Const	1		2		3		4		5			
CR	-0,01	vci46	7,88	tci25	-1,69						79	77	0,1
PH		vci46	7,30	vci40	-0,07						76	74	0,11
EX2	-0,01	tci25	-1,69	vci46	7,88						79	77	0,1
EX3	0,09	vci39	4,26	vci41	-5,27	vci44	8,93				87	84	0,08
EX4	-0,22	tci1	-3,60	vci40	7,07	vci41	-8,09	vci43	10,76		91	89	0,06
EX5	0,03	tci13	-9,38	tci14	9,15	tci19	-3,27	tci38	-5,63	tci39 7,1	94	92	0,05

Decatur

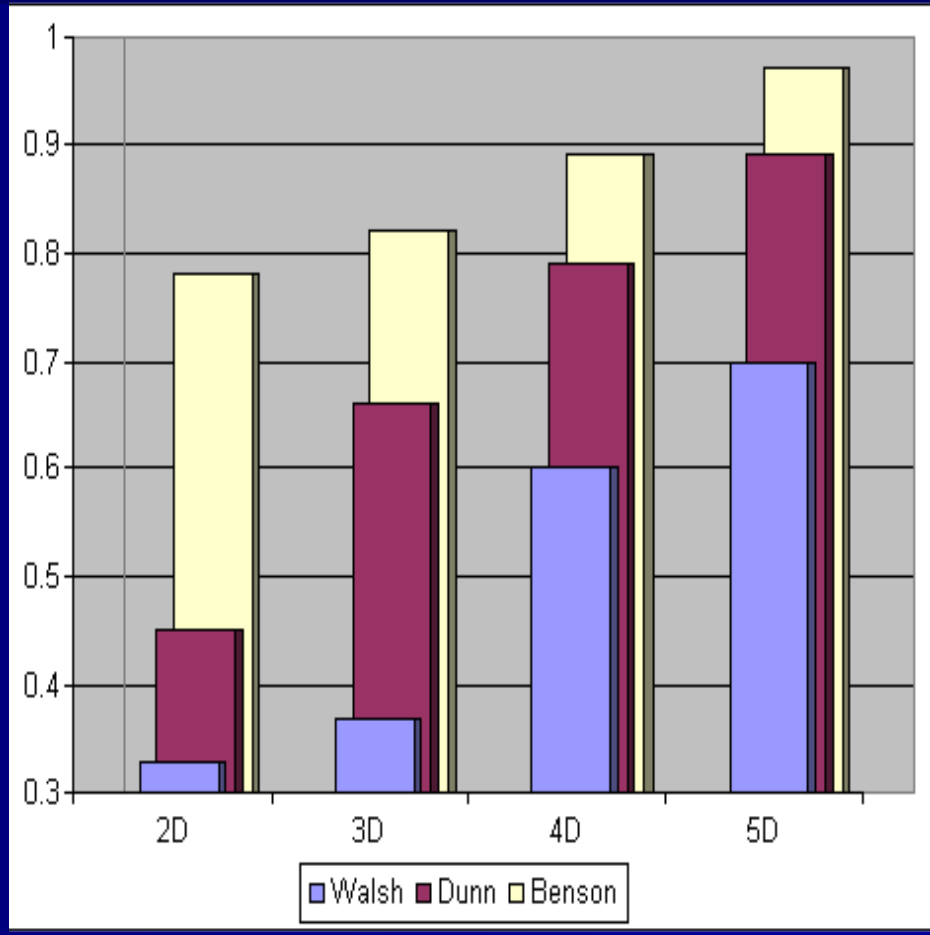
Model	Predictors and their t-crireria										D	Da	σ
	Const	1		2		3		4		5			
CR	0,00	vci43	5,25								61	58	0,15
PH		vci43	4,84	vci14	0,78						62	60	0,15
EX2	-0,01	vci45	3,65	vci46	-2,76						66	62	0,14
EX3	0,02	tci10	2,80	tci11	-2,78	vci43	6,64				74	69	0,12
EX4	0,02	tci15	-13,05	tci16	13,51	tci24	-9,47	vci17	11,08		93	92	0,06

Two Examples of 5-Factor Exhausted Models Accuracy Growth when Increasing of Predictors Amount (from 2 to 5)

Kansas

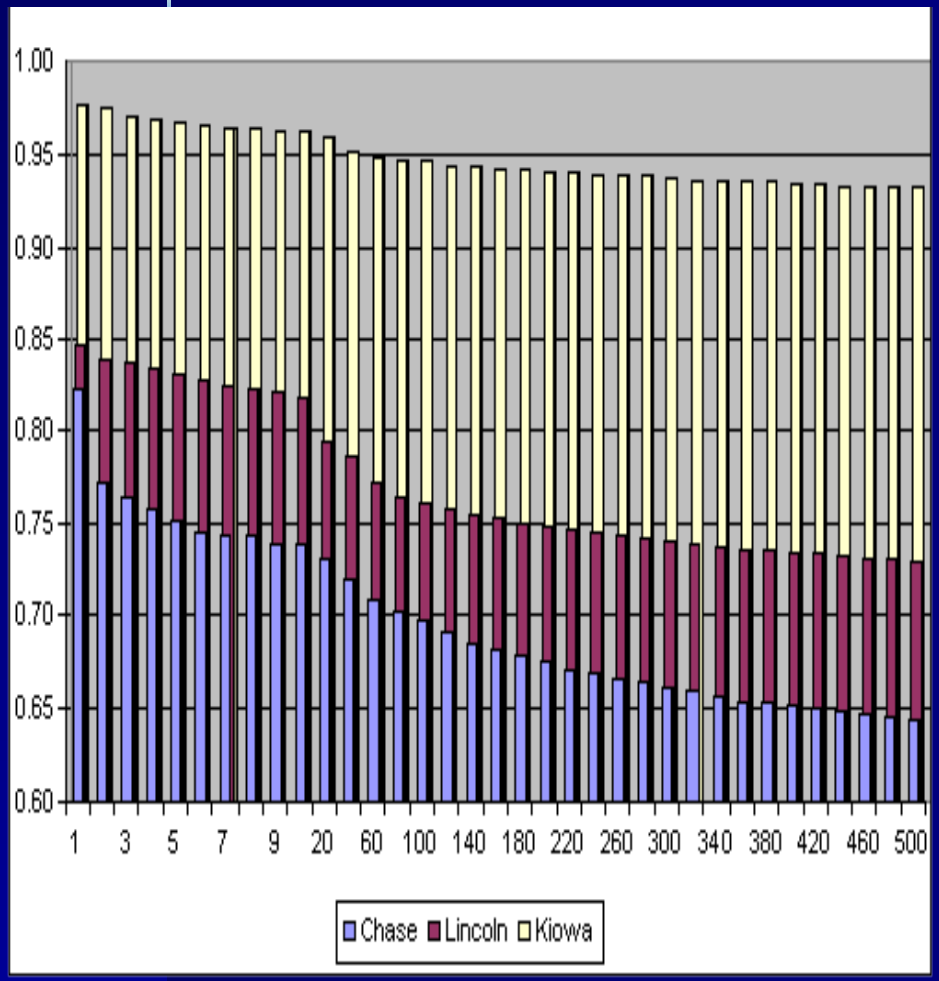


North Dakota

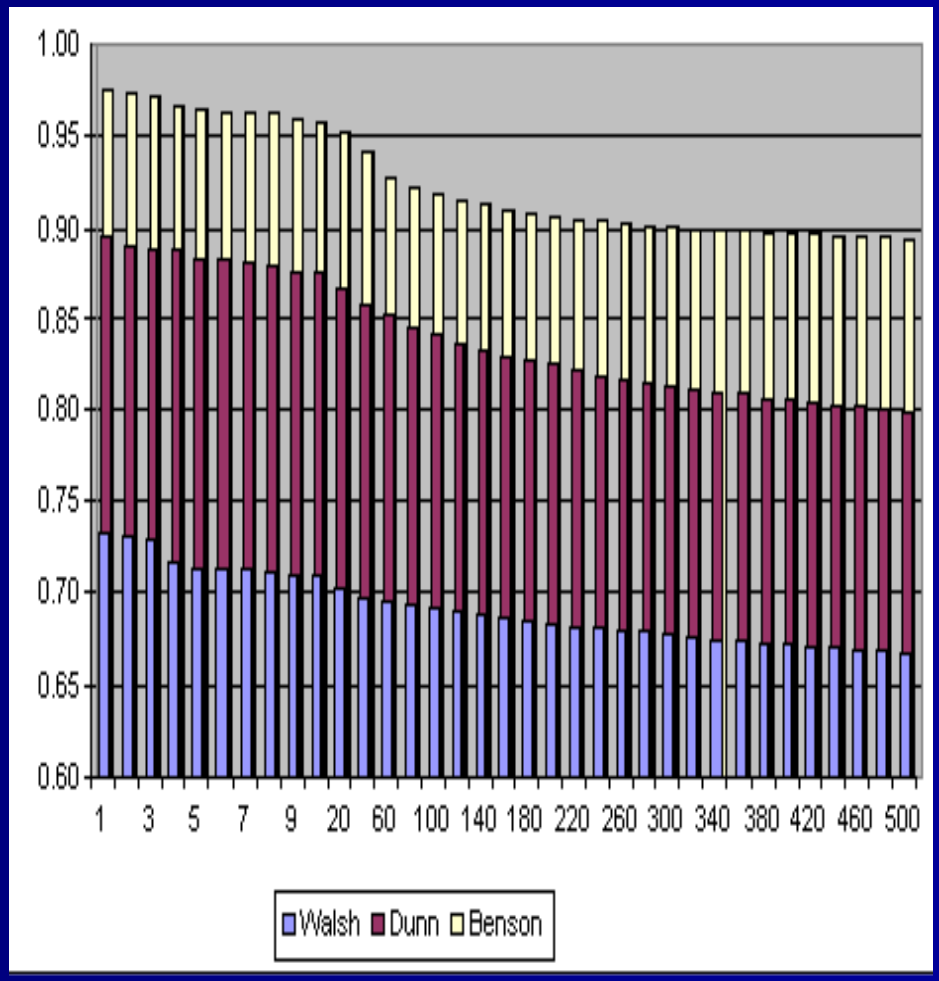


Two Examples of Determination Coefficient Decrease when Selecting the Best Exhausted 5-Factor Models (500 best models from total 100 million)

Kansas

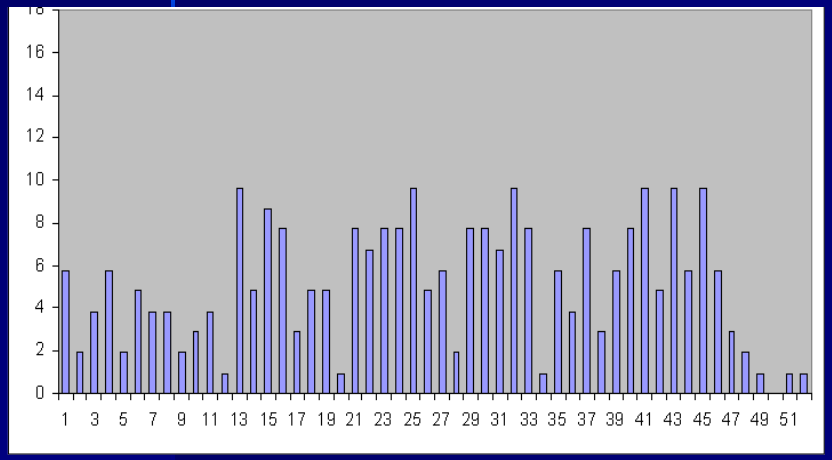


North Dakota

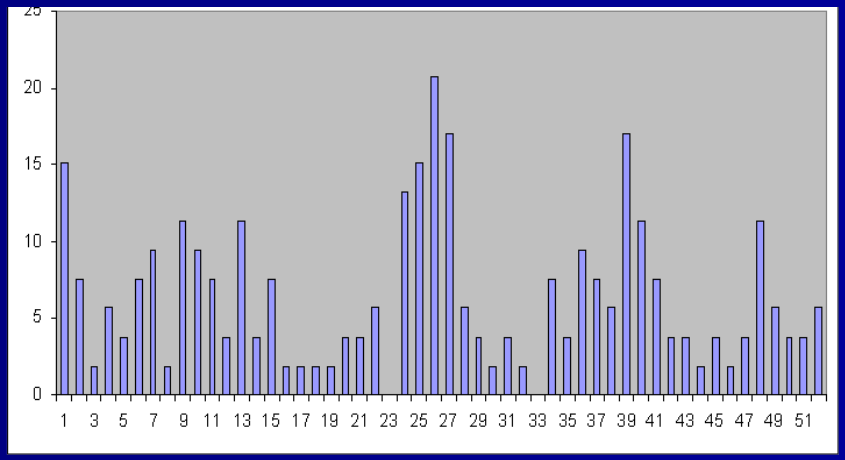


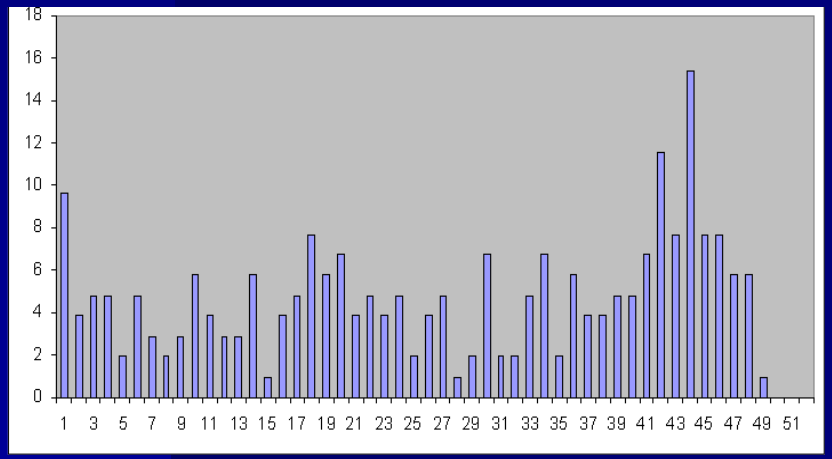
Frequency of Different Vegetation Indices Using by 5-Factor Exhausted Models

(Kansas – left, North Dakota – right)

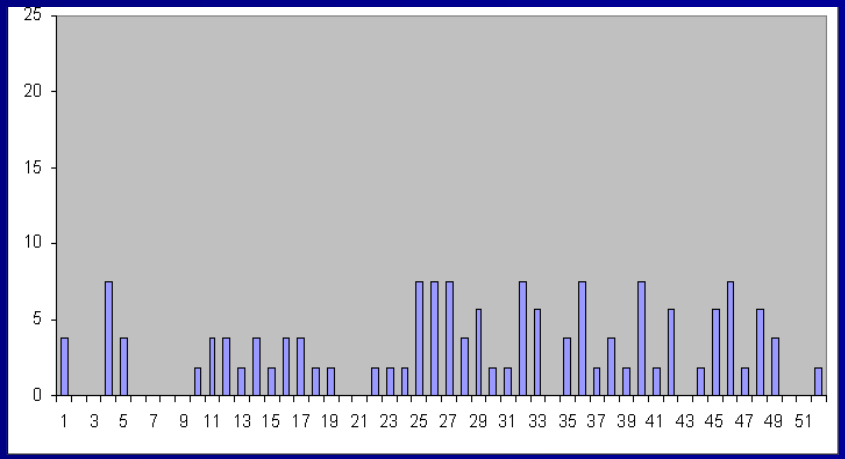


TCI





VCI



Main Conclusions

- Multifactor crop productivity regression models using as independent predictors the satellite vegetation indices can be strongly recommended to include to the modern agrometeorological forecasting techniques.

- The statistical indicators of accuracy and certainty of such models are significantly higher than the corresponding indicators of the models commonly used in agrometeorological practice especially in the case of carefully designed algorithms of its predictors selection.

- It will be very desirable if COST734 among its activity purposes set as important object the creating the all-European data banks of long-term satellite vegetation indices as well as the historical series of the most important crop production and areas for small administrative regions of European countries.

Many Thanks for Your Attention