Predicting West Nile Virus Incidence in ND using Machine Learning Techniques

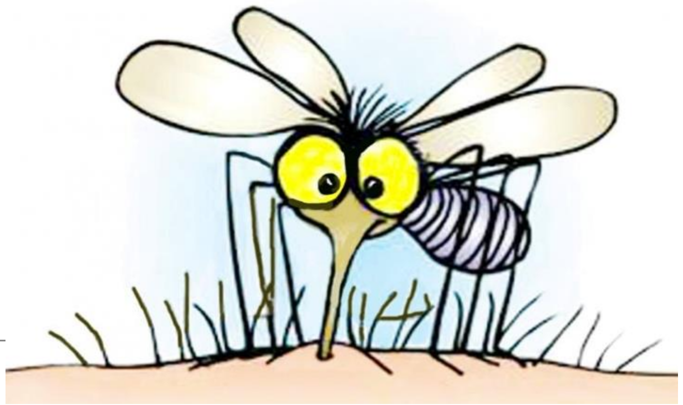
PROJECT PARTICIPANTS

1. MITCH CAMPION, M.S. GRADUATE STUDENT
2. MARTIN POZNIAK, UNDERGRADUATE EE STUDENT.
3. PRAKASH RANGANATHAN, PH.D. ASSISTANT PROFESSOR (PI)
4. NAIMA KAABOUCH, PH.D. ASSOCIATE PROFESSOR (CO-PI)

NDSU

1. MARK BOETL, ENTOMOLOGY PROFESSOR

Project Objectives



- GOALS
- Develop an user Interface showing history of North Dakota cases on West Nile Virus and Culex Species Type by categorize risk levels
  - *The interface utilizes multilayered Google Maps developed through Google Fusion Tables*
  - *An understanding of historical data and weather variables is essential for providing sufficient lead time to predict WNV occurrence, and for implementing disease control and prevention strategies such as spray period and hiring of seasonal mosquito workers.*
- Develop a Forecasting Model to predict mosquito counts and WNV Incidence using Weather, and prior years of mosquito trap count data.

- NASA Relevance
- Use Remote sensing, weather data sets, “Better Understand Earth”

Methodology

Data Collection Data Organization

- Receive data from county mosquito control, NDDoH
- Integrate Weather data
- Organize data to useful format
- Clean data etc..

Analysis and Forecasting

- Basic statistical examination
- Feature selection
- Forecasting algorithms
- Accuracy assessment

Data Visualization Analysis Web Integration

- Easy interactive web interface
- Basic analysis

Decision Protocol Management

- Decision-making protocols
  - hiring of seasonal mosquito control units
  - public health awareness campaigns

Data Collection, Aggregation

- Data must be organized to be used and interpreted
  - Data comes from many sources in many formats making for a time consuming challenge

Weekly Mosquito Trap Counts

Dates: July 1st - July 7th 2013

Counties	Trap Location	Male	Female									Total Mosquitoes
			Anop-heles	Aedes	Aedes vexans	Culex	Culex Tarsalis	Culex salinarius	Culiseta	Other	Total Female	
Region I												
Williams	Williston #1	0	0	0	2	3	0	0	0	0	5	5
Williams	Williston #2	80	0	48	104	16	0	0	72	0	240	320
Williams	Tioga										0	0
Divide	Crosby	96	0	80	64	24	0	0	56	0	224	320
McKenzie	Watford City										0	0
Mountrail	Stanley										0	0
Region I Total		**176**	**0**	**128**	**170**	**43**	**0**	**0**	**128**	**0**	**469**	645
Region II												
Ward	Minot Oak Park(1)	12	0	20	0	0	0	0	3	0	23	35
Ward	Minot NW (2)	2	0	16	0	0	0	0	11	0	27	29
Ward	Ryder	8	0	14	6	0	4	0	4	0	28	36
Bottineau	Bottineau	6	2	13	1	0	0	0	1	0	17	23
Burke	Bowbells	40	4	96	44	8	0	0	12	0	164	204
McHenry	Towner	8	0	12	4	4	0	0	0	0	20	28
McLean	Deep Wtr Creek Bay	TWTC									0	0
McLean	Washburn	8	0	12	0	4	0	0	0	0	16	24
Renville	Mohall	80	32	224	32	0	16	0	16	0	320	400
Sheridan	Martin										0	0
Region II Total		**164**	**38**	**407**	**87**	**16**	**20**	**0**	**47**	**0**	615	779
Region III												
Ramsey	Devils Lake #1	4	0	20	22	10	2	0	10	0	64	68
Ramsey	Devils Lake #2										0	0

Sheet1 | Sheet2 | Sheet3

Sample trap count data set from ND DoH

Aggregation

Counties	Trap.Location	Male	Anop.heles	Aedes	Aedes.vexans	Culex	Culex.Tarsalis	Culex.sali	Culiseta	Other	Total.Female	Total.Mosqui	Week.Start	Week.End	Year	Week.	MaxTemp	MinTemp	AvgTemp	Rainfall	AvgRain	MaxHumidi	MinHumid	AvgHumid	WindDir	AvgWind	MaxWind	MeanMaxTemp	MeanMaxHumidity	avgCT	jantemp	febtemp	marchter	apriltemp	maytemp
Grand Fork	Grand Forks #1	136	8	112	80	0	0	0	72	0	272	408	5/31/2005	6/6/2005	2005	1	89	51	70.28571	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.71428571	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	23	1	17	18	0	0	0	41	0	77	100	5/31/2005	6/6/2005	2005	1	89	51	70.28571	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.71428571	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	108	10	164	108	0	0	0	152	0	434	542	6/7/2005	6/12/2005	2005	2	76	50	63	2.7	0.45	100	46	78.1667	158.7	10.333	38	71.33333333	98.16666667	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	10	2	58	40	0	0	0	58	0	158	168	6/7/2005	6/12/2005	2005	2	76	50	63	2.7	0.45	100	46	78.1667	158.7	10.333	38	71.33333333	98.16666667	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	30	0	56	64	0	0	0	8	0	128	158	6/13/2005	6/19/2005	2005	3	88	52	67.42857	0.78	0.1114	100	41	75.7143	158.6	9.2857	30	76.14285714	92.71428571	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	6	0	82	8	0	0	0	12	0	102	108	6/13/2005	6/19/2005	2005	3	88	52	67.42857	0.78	0.1114	100	41	75.7143	158.6	9.2857	30	76.14285714	92.71428571	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	176	0	576	160	64	0	0	48	0	848	1024	6/20/2005	6/25/2005	2005	4	93	51	72	0.15	0.025	100	32	68.1667	202.3	8.8333	26	84.66666667	92.5	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	312	8	552	144	24	0	0	40	0	768	1080	6/20/2005	6/25/2005	2005	4	93	51	72	0.15	0.025	100	32	68.1667	202.3	8.8333	26	84.66666667	92.5	0	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	88	0	0	80	0	40	0	8	0	488	576	6/26/2005	7/2/2005	2005	5	85	48	66.14286	2.96	0.4229	100	47	80.2857	186.3	11	40	75	97.14285714	22	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	48	0	140	116	20	4	0	12	0	292	340	6/26/2005	7/2/2005	2005	5	85	48	66.14286	2.96	0.4229	100	47	80.2857	186.3	11	40	75	97.14285714	22	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	0	1	46	7	22	38	0	0	0	114	114	7/4/2005	7/10/2005	2005	6	93	48	72.14286	0.17	0.0243	100	49	75.2857	190.1	10.714	33	82.14285714	94.28571429	27	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #2	44	0	216	152	96	16	0	12	0	492	536	7/4/2005	7/10/2005	2005	6	93	48	72.14286	0.17	0.0243	100	49	75.2857	190.1	10.714	33	82.14285714	94.28571429	27	11.194	21.5	33.226	59.3	63.161
Grand Fork	Grand Forks #1	0	0	132	32	0	0	0	0	0	164	164	7/11/2005	7/17/2005	2005	7	90	58	74.42857	0.2	0.0286	100	34	70.4286	248.4	7.1429	28	86.14285714	93.28571429	6	11.194	21.5	33.226	59.3	63.161

Subset large dataset into weekly subsets
(week 1 subset shown below)

Counties	Trap.Location	Male	Anop.heles	Aedes	Aedes.vexans	Culex	Culex.Tarsalis	Culex.sali	Culiseta	Other	Total.Fem	Total.Mosc	Week.Start	Week.End	Year	Week.	MaxTemp	MinTemp	AvgTemp	Rainfall	AvgRain	MaxHumidi	MinHumid	AvgHumid	WindDir	AvgWind	MaxWind	MeanMaxTemp	MeanMaxHum	avgCT	jantemp	febtemp	marchtemp	apriltemp	maytemp	MayAprRain	MayRain
Grand Fo	Grand Forks #1	0	0	0	0	0	0	0	0	0	0	0	5/29/2006	6/2/2006	2006	1	89	51	70.2857	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.7142857	0	28.484	16.25	30.16129	60.367	68.129	3.43	2.4
Grand Fo	Grand Forks #2	0	0	0	0	0	0	0	0	0	0	0	5/29/2006	6/2/2006	2006	1	89	51	70.2857	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.7142857	0	28.484	16.25	30.16129	60.367	68.129	3.43	2.4
Grand Fo	Grand Forks #1	0	2	25	7	30	17	0	5	1	87	87	5/28/2007	6/4/2007	2007	1	85	42	64.5	0.6	0.075	100	37	72	229.6	11.625	31	76	93.125	12.5	20.387	12.321	35.612903	53.4	69.065	5.7	5.17
Grand Fo	Grand Forks #2	0	0	22	39	3	8	0	0	0	72	72	5/28/2007	6/4/2007	2007	1	85	42	64.5	0.6	0.075	100	37	72	229.6	11.625	31	76	93.125	12.5	20.387	12.321	35.612903	53.4	69.065	5.7	5.17
Grand Fo	Grand Forks #1	0	0	2	6	0	0	0	0	0	8	8	5/26/2008	6/2/2008	2008	1	82	27	57.125	0.56	0.07	100	18	59.125	131.8	9	26	70.75	85.625	0	13.161	14.931	29.677419	52.3	66.097	1.66	1.06
Grand Fo	Grand Forks #2	0	0	2	0	0	0	0	1	0	3	3	5/26/2008	6/2/2008	2008	1	82	27	57.125	0.56	0.07	100	18	59.125	131.8	9	26	70.75	85.625	0	13.161	14.931	29.677419	52.3	66.097	1.66	1.06
Grand Fo	Grand Forks #1	0	0	40	0	2	0	0	0	0	42	42	5/25/2009	6/1/2009	2009	1	79	38	55.875	0.83	0.1038	100	19	67.625	169.5	10.875	33	69.125	93.625	0	10	17.679	29.129032	48.533	63.871	2.7	1.39
Grand Fo	Grand Forks #2	0	0	4	0	1	0	0	0	0	5	5	5/25/2009	6/1/2009	2009	1	79	38	55.875	0.83	0.1038	100	19	67.625	169.5	10.875	33	69.125	93.625	0	10	17.679	29.129032	48.533	63.871	2.7	1.39
Grand Fo	Grand Forks #1	0	0	36	7	1	2	0	35	0	81	81	5/31/2010	6/6/2010	2010	1	78	43	61	0.41	0.0586	100	38	68.1429	269.1	7.5714	26	72.85714286	93.1428571	1	14.806	17.643	39.967742	62.1	66.968	5.95	4.61
Grand Fo	Grand Forks #2	0	0	31	0	0	0	0	17	0	48	48	5/31/2010	6/6/2010	2010	1	78	43	61	0.41	0.0586	100	38	68.1429	269.1	7.5714	26	72.85714286	93.1428571	1	14.806	17.643	39.967742	62.1	66.968	5.95	4.61
Grand Fo	Grand Forks #1	0	0	8	0	0	0	0	12	0	20	20	5/30/2011	6/5/2011	2011	1	84	46	62.2857	1.17	0.1671	100	23	66.8571	213.7	13.714	36	72.85714286	86	0	10	18.607	27.645161	50.367	63.903	4.95	2.5
Grand Fo	Grand Forks #2	0	0	3	0	0	0	0	14	0	17	17	5/30/2011	6/5/2011	2011	1	84	46	62.2857	1.17	0.1671	100	23	66.8571	213.7	13.714	36	72.85714286	86	0	10	18.607	27.645161	50.367	63.903	4.95	2.5
Grand Fo	Grand Forks #1	0	0	0	0	0	0	0	0	0	0	0	5/28/2012	6/3/2012	2012	1	81	33	55.8571	0.12	0.0171	100	25	68.7143	263.3	7.2857	25	68.85714286	94.5714286	0	26.194	29.31	49.354839	58.667	70.581	3.28	1.77
Grand Fo	Grand Forks #2	0	0	0	0	0	0	0	0	0	0	0	5/28/2012	6/3/2012	2012	1	81	33	55.8571	0.12	0.0171	100	25	68.7143	263.3	7.2857	25	68.85714286	94.5714286	0	26.194	29.31	49.354839	58.667	70.581	3.28	1.77
Grand Fo	Grand Forks #1	0	0	0	0	0	0	0	0	0	0	0	5/27/2013	6/2/2013	2013	1	77	40	60.7143	1.15	0.1643	100	39	73.2857	141.1	11.286	30	69.42857143	92	0	19	19.679	24.935484	39.5	65.871	6.47	4.83
Grand Fo	Grand Forks #2	0	0	0	0	0	0	0	0	0	0	0	5/27/2013	6/2/2013	2013	1	77	40	60.7143	1.15	0.1643	100	39	73.2857	141.1	11.286	30	69.42857143	92	0	19	19.679	24.935484	39.5	65.871	6.47	4.83
Grand Fo	Grand Forks #1	0	0	0	0	0	0	0	0	0	0	0	5/26/2014	6/1/2014	2014	1	89	51	70.2857	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.7142857	0	13.806	11	27.903226	47.267	65.839	5.3	2.61
Grand Fo	Grand Forks #2	0	0	0	0	0	0	0	0	0	0	0	5/26/2014	6/1/2014	2014	1	89	51	70.2857	0.6	0.0857	100	21	68.4286	89.86	7.7143	28	82.57142857	93.7142857	0	13.806	11	27.903226	47.267	65.839	5.3	2.61
Grand Fo	Grand Forks #1	9	0	0	35	8	0	NA	11	8	62	71	5/25/2015	5/29/2015	2015	1	75	89.8	85	38	62.8	0.66	0.132	100	23	66.8	283.2	10.8	31	1	22.065	14.679	42.935484	58.967	65.516	5.27	4.47
Grand Fo	Grand Forks #2	10	0	0	9	14	2	NA	37	3	65	75	5/25/2015	5/29/2015	2015	1	75	89.8	85	38	62.8	0.66	0.132	100	23	66.8	283.2	10.8	31	1	22.065	14.679	42.935484	58.967	65.516	5.27	4.47

Trap Count Forecasting Model – PLSR

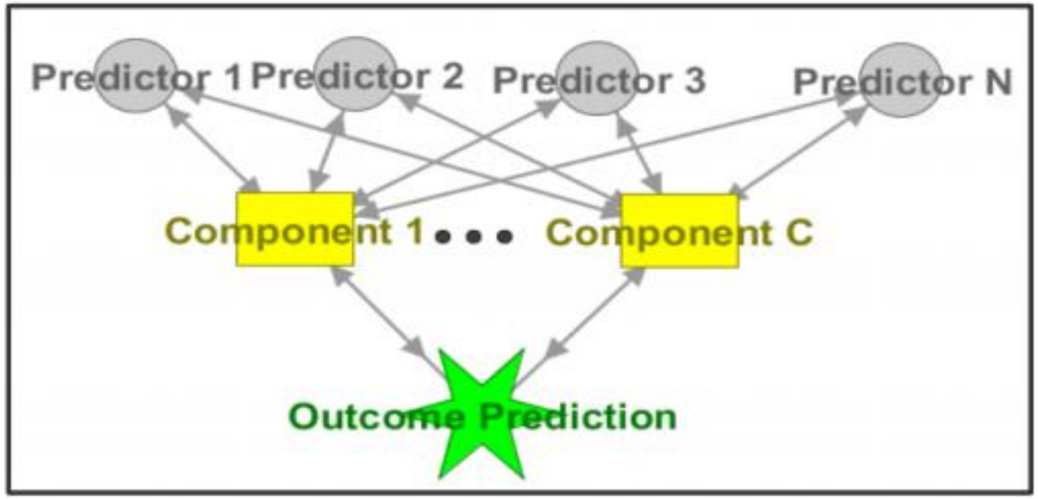


Fig. 1. Block Diagram of partial least squares algorithm

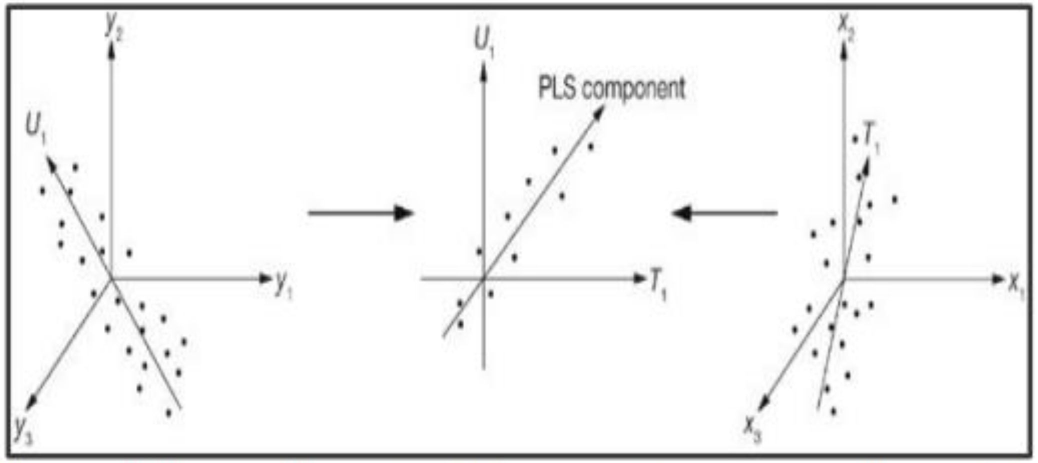


Fig. 2. Predictor and target variable spaces forming maximal covariant regression

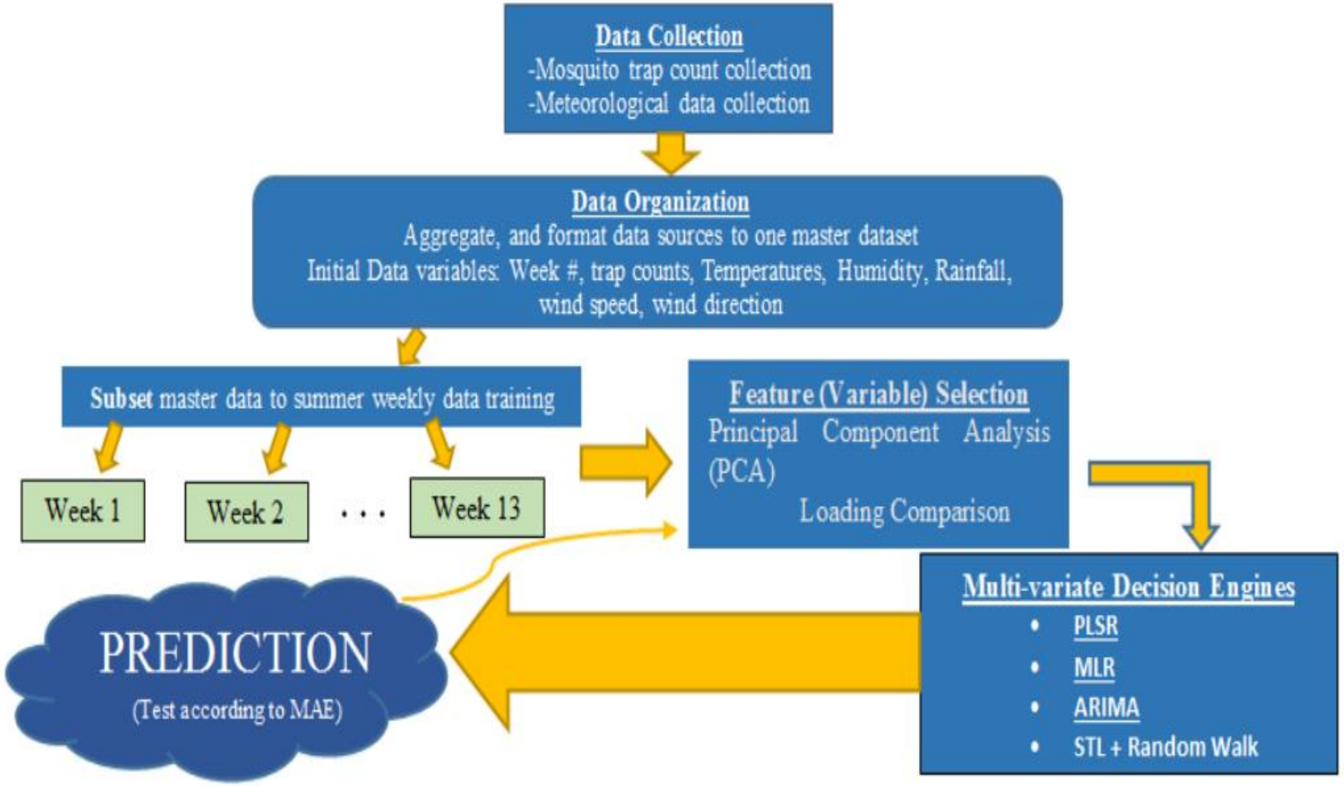


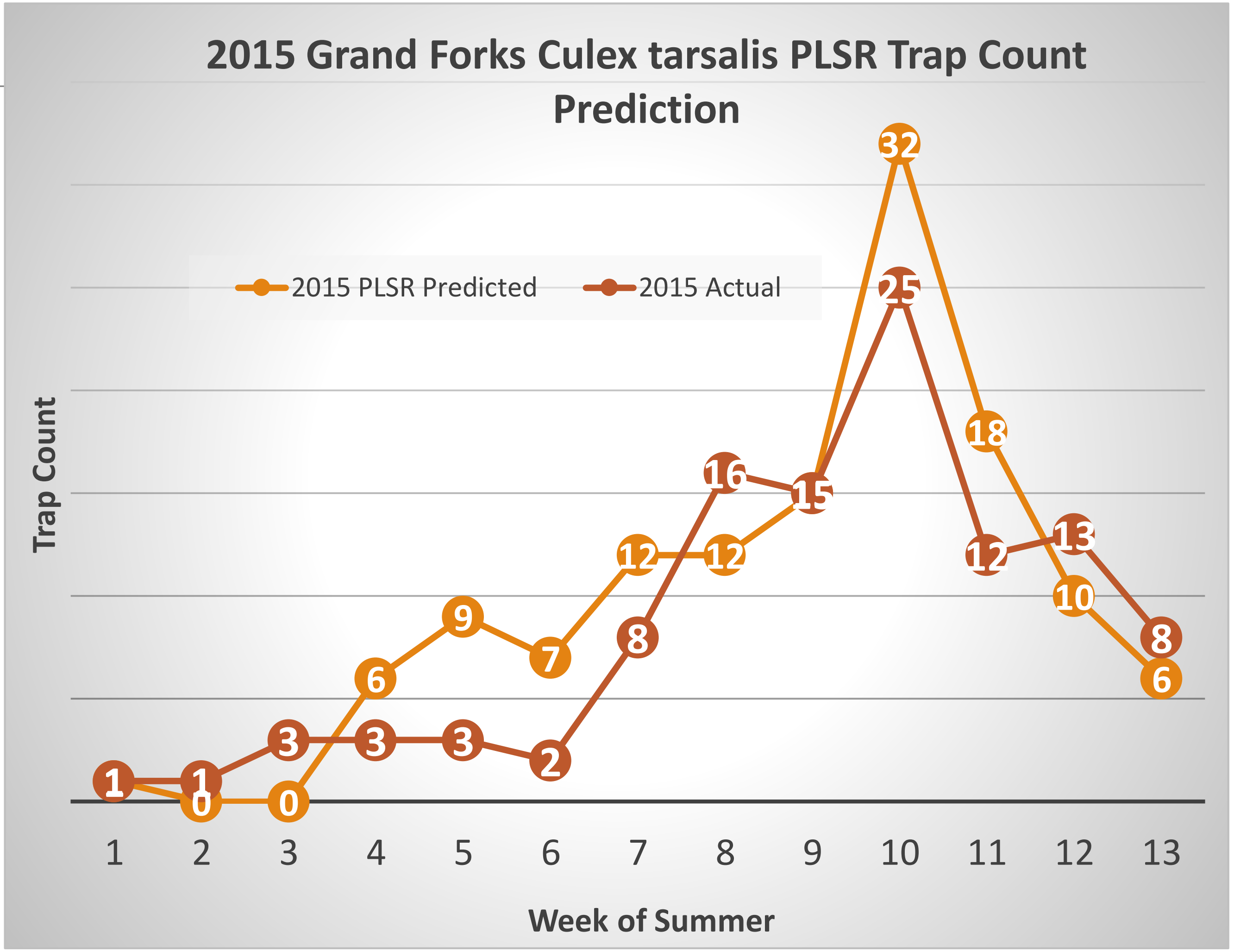
Fig. 6. Block diagram showing start to finish modeling process

Prediction Results

TABLE I. TABULAR DISPLAY OF DIFFERENT PREDICTION ALGORITHMS COMPARED WITH THE ACTUAL TRAP COUNT FOR 2015 AND THE CORRESPONDING MAE OF EACH PREDICTION. PLSR SHOWS THE LEAST AMOUNT OF PREDICTION ERROR

2015 Trap Count Prediction Model Comparison

Week	2015 Actual	STL+RandomWalk	Holt-Winters	ARIMA	PLSR
1	1	0	0	0	1
2	1	0	1	0	0
3	3	0	1	13	0
4	3	0	3	0	6
5	3	6	2	6	9
6	2	0	12	13	7
7	8	1	16	13	12
8	16	26	16	9	12
9	15	8	2	18	15
10	25	40	15	16	32
MAE		*5.62*	*4.5*	*5.3*	*3.3*



WNV Human Cases Binomial Prediction Modeling

- Using binomial prediction methods, Random Forest, Regression tree, logistic regression
- Predict based on meteorological variables and trap count whether WNV will be contracted by a human

Random Forest Prediction confusion matrix results

	RandomForest		Error
Actual	0	1	
0	260	0	0
1	2	8	0.2

Risk Maps Website – Google Fusion Tables

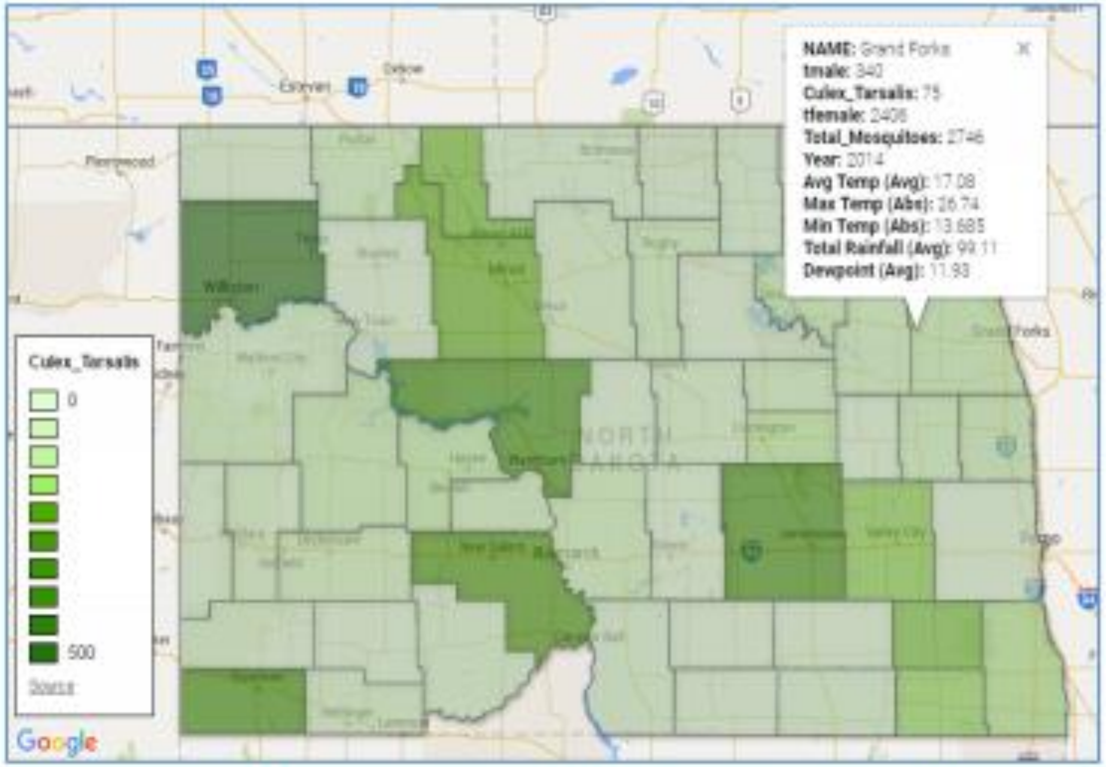


Fig. 4. Layered Google Map Showing Culex Tarsalis in colored layers and a pop up box displaying Grand Forks County information

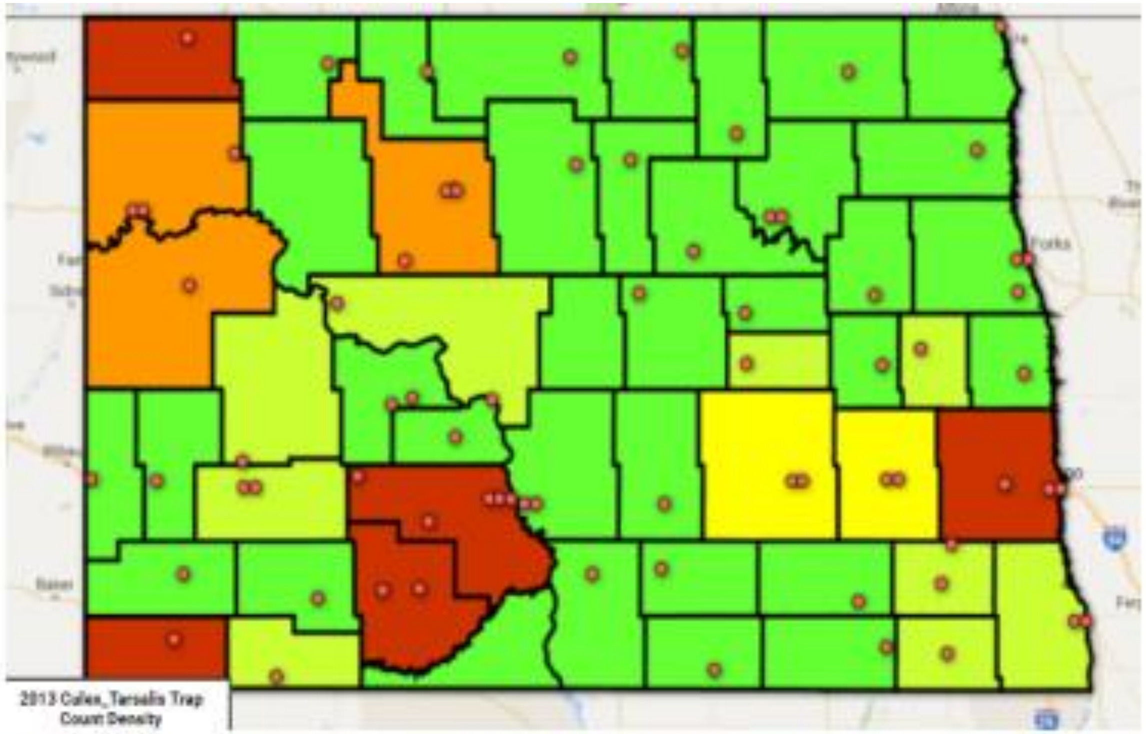
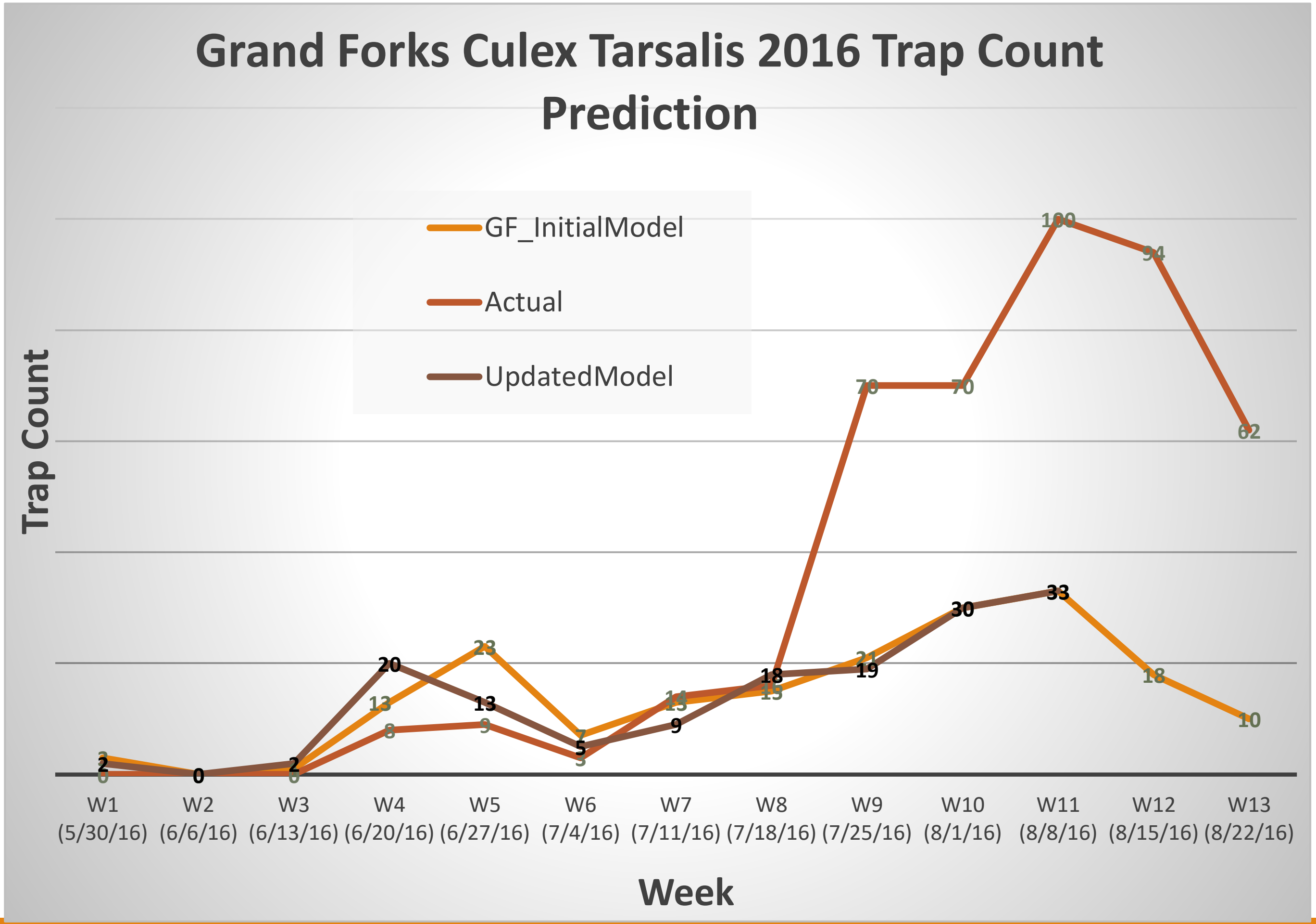


Fig. 3. Layered Google Map displaying colors that correspond to mosquito trap counts. Red counties indicate high population, green indicate low. Red markers correspond to actual trap locations in North Dakota. This type of map will be used to show trap counts and WNV risk

[WEBSITE]

Future Work

- Train student or department to do modeling
- Improve Accuracy of Culex tarsalis trap count model
- Continued development and accuracy improvement of human WNV contraction model
  - Further analysis of Minimum Infection Rate (MIR) data
  - Vector Index (VI)
- Extend the scope of prediction models to all counties
- Collection and addition of key weather variables to website analysis interface



Publications

M. Campion, C. Bina, **P. Ranganathan** et.al. Predicting West Nile Virus (WNV) Occurrences in North Dakota using Data Mining Techniques FTC 2016 - Future Technologies Conference 2016, 6-7 December 2016 | San Francisco, United States.

- Planning to submit new additional findings to “Journal of Entomology”

FTC 2016 - Future Technologies Conference 2016
6-7 December 2016 | San Francisco, United States

Predicting West Nile Virus (WNV) Occurrences in North Dakota using Data Mining Techniques

[1]Mitch Campion, [1]Calvin Bina, [1]Martin Pozniak, [2]Todd Hanson, [3]Jeff Vaughan, [4]Joseph Mehus, [5]Scott Hanson, [6]Laura Cronquist, [6]Michelle Feist, [1]Prakash Ranganathan, [1]Naima Kaabouch, [7] Mark Boetel
[1]Department of Electrical Engineering, University of North Dakota, Grand Forks, ND 58202
mitchell.campion@ndus.edu