Statistical Perspectives On The Human Factor from RBN and WSPR Networks

David Campbell
Dr. Deborah Kunkel

The Data

- Data from RBN and WSPR networks
  - Includes “Spots”
    - Spots include Time, Date, Callsign Receiving and Broadcasting, Locations, Frequency, Band
- This is the raw data we got for F5IN callsign:

	X	callsign	de_pfx	de_cont	freq	band	dx	dx_pfx	dx_cont	mode	db	date	speed
1	1	DF7GB	DL	EU	3511.0	80m	F5IN	FALSE	EU	CQ	11	2016-01-01 06:22:12	21
2	2	SV8RV	SV	EU	7004.1	40m	F5IN	FALSE	EU	CQ	15	2016-01-01 06:39:01	24
3	3	K3LR	K	<NA>	7004.0	40m	F5IN	FALSE	EU	CQ	22	2016-01-01 06:39:03	23
4	4	W1NT	K	<NA>	7004.1	40m	F5IN	FALSE	EU	CQ	15	2016-01-01 06:39:04	24
5	5	HA2KSD	HA	EU	7004.0	40m	F5IN	FALSE	EU	CQ	35	2016-01-01 06:39:07	24
6	6	N2QT	K	<NA>	7004.0	40m	F5IN	FALSE	EU	CQ	7	2016-01-01 06:39:07	23

	tx_mode	mdhead_callsign	mdhead_dx
1	CW	JN49CX	JN18DD
2	CW	KM07KS	JN18DD
3	CW	EN91SE	JN18DD
4	CW	FN32PC	JN18DD
5	CW	JN87WB	JN18DD
6	CW	FM07II	JN18DD

The Data

- Cleaning the Data
  - Focused on the data for our model and rescaled all the variables

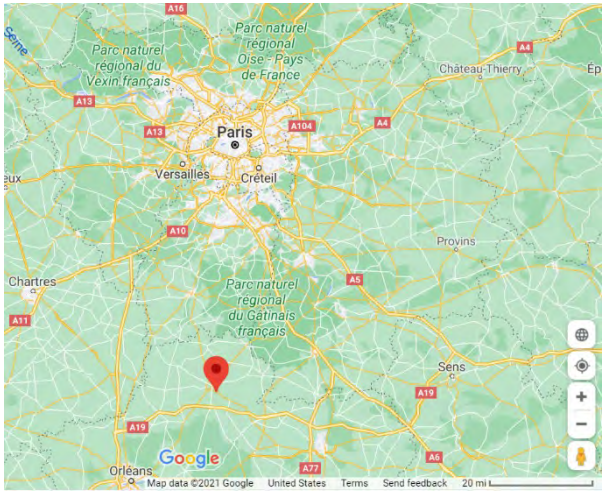
	reporter	reporter.grid	callsign	trans.grid	distance	utc	utc.num	reporter.lat
1	VE6JY	DO33or	05JVD	BM39	3398	2021-02-09 03:24:00	204	53.02083
2	VE6JY	DO33or	2E0DLC	IO93hm	6525	2021-02-09 10:38:00	638	53.02083
3	VE6JY	DO33or	2E0DLC	IO93hm	6525	2021-02-09 10:20:00	620	53.02083
4	VE6JY	DO33or	2E0DLC	IO93hm	6525	2021-02-09 10:02:00	602	53.02083
5	VE6JY	DO33or	2E0DSS	IO82xl	6585	2021-02-09 16:16:00	976	53.02083
6	VE6JY	DO33or	2E0DSS	IO82xl	6585	2021-02-09 15:40:00	940	53.02083

	reporter.long	date	no_from_callsign_date	total.location	total.date	total.date.and.location
1	-113.9583	2017-12-28	1	1	1089424	1
2	-113.9583	2017-12-13	2	15067	874296	731
3	-113.9583	2017-12-24	5	15067	1222007	323
4	-113.9583	2017-12-26	4	15067	1173931	1137
5	-113.9583	2017-12-01	2	25031	1042469	682
6	-113.9583	2017-12-08	1	25031	946949	462

	rescaled.total.date	rescaled.total.location	rescaled.total.date.and.location	rescaled.distance
1	0.5791597	-0.59511266	-0.69446027	-0.07921069
2	-1.2296018	-0.27460412	-0.28887144	1.01116853
3	1.6938962	-0.27460412	-0.51555671	1.01116853
4	1.2896809	-0.27460412	-0.06329738	1.01116853
5	0.1843696	-0.06263366	-0.31609590	1.03209042
6	-0.6187471	-0.06263366	-0.43832815	1.03209042

Spot Watcher

- We have created an app to help with visualizations of the use of certain spots
  - https://Deborah-kunkel.shinyapps.io/SpotWatcher/
  - This app focuses on Callsign F5IN: maidenhead coordinate JN18dd (Dadonville, France)



Statistical Model

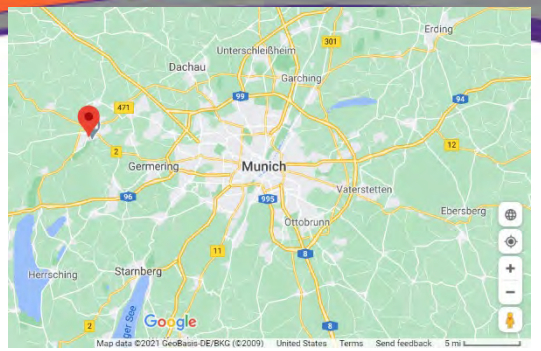
- Motivation
  - Spikes in the counts of spots
  - Goal: estimate a function that predicts the number of spots heard from a fixed location on a particular date.
- Poisson Generalized Linear Model (GLM)
  - Why this model?
    - Counts had a similar distribution to the Poisson Distribution
  - Response Variable: Count of spots to a callsign on a date
  - Predictor Variables:
    - Count of spots on a date
    - Count of spots from each location
    - Count of spots from each location on a date
    - Distance to callsign
      - Used a spline for this variable to use a flexible function to model the effect of distance

CLEMSON UNIVERSITY

Statistical Model

- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}{}^2 + \beta_5\, g(d(s))$
- $\lambda(t,s)$: Function of counts of spots that a callsign receives given date, t, and location, s.
- $X_{1,t}$:Number of spots from date, t
- $X_{2,s}$:Number of spots from location, s
- $X_{3,t,s}$:Number of spots from date, t, and location, s
- $g(d(s))$:Smooth function of distance from receiver to transmitter
- To create this model, we used data from the WSPR network for December 2017

CLEMSON UNIVERSITY



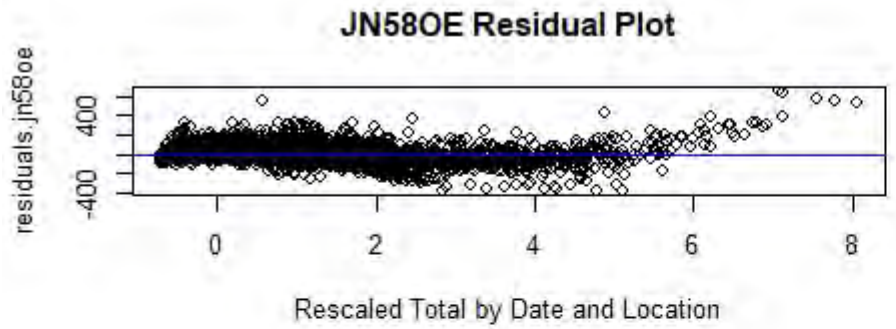
Final Model and Results

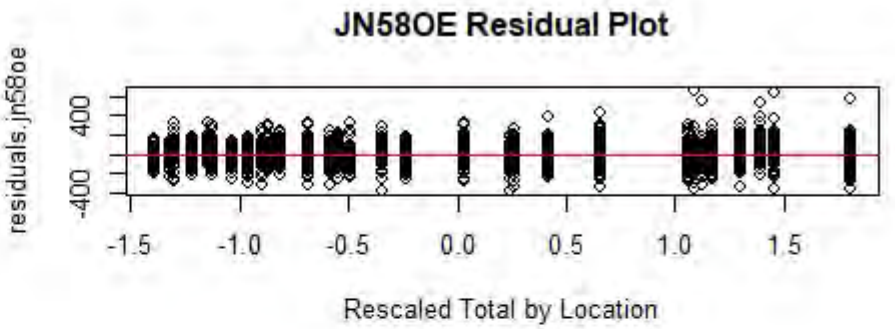
- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}{}^2 + \beta_5\, g(d(s))$
- Using Data for Callsign DK8FT (maidenhead JN58oe in Fürstenfeldbruck, Germany):

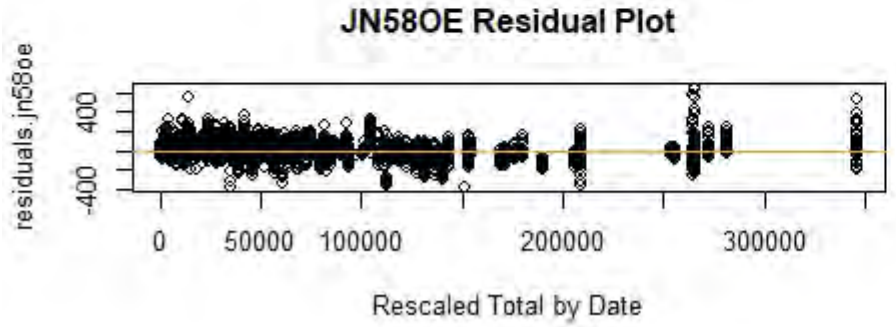
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.262404	0.009477	555.312	< 2e-16	***
poisson.data.rescaled.total.date	0.062412	0.001163	53.657	< 2e-16	***
poisson.data.rescaled.total.location	-0.019191	0.001332	-14.411	< 2e-16	***
poisson.data.rescaled.total.date.and.location.2	-0.116739	0.000364	-320.726	< 2e-16	***
poisson.data.rescaled.total.date.and.location	1.041081	0.001885	552.233	< 2e-16	***
X1	-2.659388	0.026629	-99.867	< 2e-16	***
X2	-2.226499	0.017113	-130.107	< 2e-16	***
X3	-1.971727	0.015888	-124.102	< 2e-16	***
X4	-1.677650	0.011921	-140.725	< 2e-16	***
X5	-1.077331	0.013184	-81.713	< 2e-16	***
X6	-1.465921	0.013806	-106.182	< 2e-16	***
X7	-1.355001	0.012026	-112.672	< 2e-16	***
X8	-1.072345	0.011784	-91.000	< 2e-16	***
X9	-1.306203	0.011370	-114.884	< 2e-16	***
X10	-1.496752	0.011509	-130.055	< 2e-16	***
X11	-1.356888	0.011641	-116.560	< 2e-16	***
X12	-1.179067	0.024574	-47.979	< 2e-16	***
X13	-5.907961	0.054968	-107.479	< 2e-16	***
X14	-0.394943	0.127750	-3.092	0.00199	**
X15	-6.719685	0.185455	-36.233	< 2e-16	***

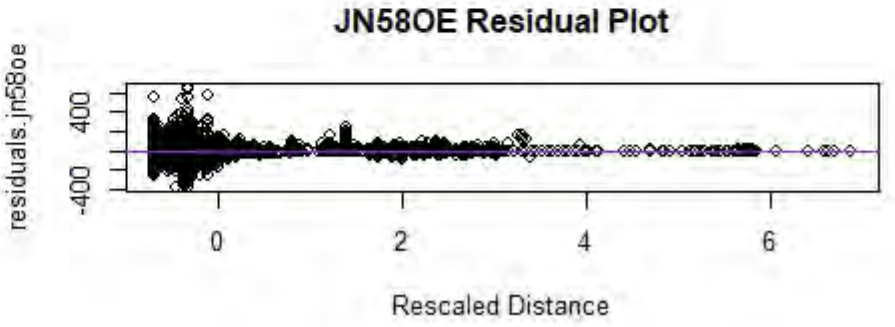
Final Model and Results

- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}{}^2 + \beta_5\, g(d(s))$

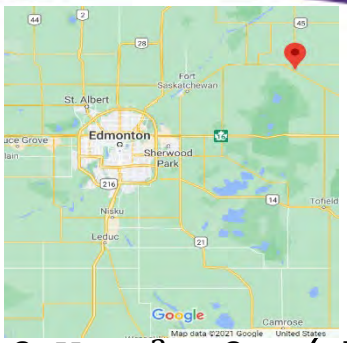








Final Model and Results

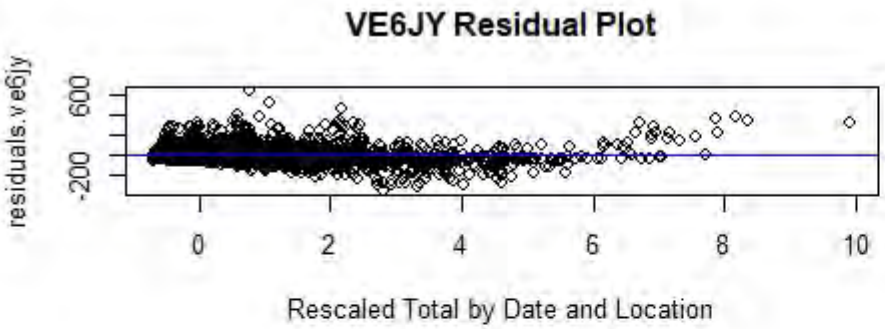


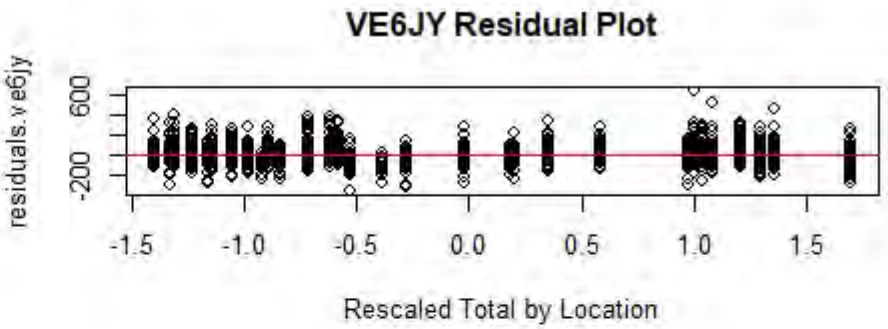
- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}^2 + \beta_5\, g(d(s))$
- Using data for callsign VE6JY (maidenhead DO33or in Lamont, Alberta, Canada):

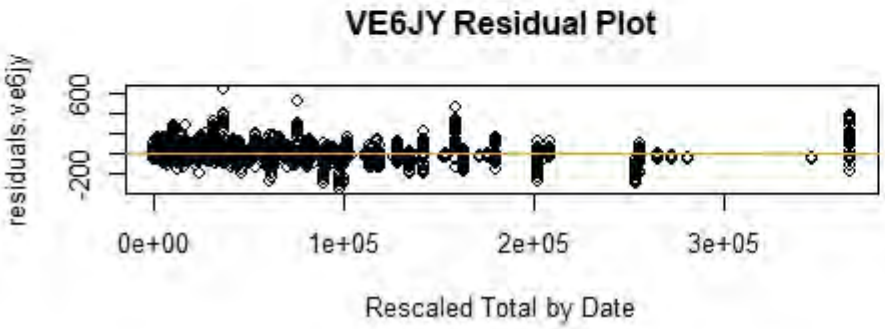
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	4.675945	0.020551	227.53	<2e-16	***
poisson.data.rescaled.total.date	0.093745	0.001526	61.44	<2e-16	***
poisson.data.rescaled.total.location	0.082056	0.002013	40.77	<2e-16	***
poisson.data.rescaled.total.date.and.location.2	-0.120541	0.000483	-249.57	<2e-16	***
poisson.data.rescaled.total.date.and.location	0.991939	0.002549	389.12	<2e-16	***
X1	-2.218985	0.046390	-47.83	<2e-16	***
X2	0.397394	0.027302	14.55	<2e-16	***
X3	-0.833218	0.024946	-33.40	<2e-16	***
X4	-0.307937	0.021654	-14.22	<2e-16	***
X5	-0.833561	0.023143	-36.02	<2e-16	***
X6	-0.961769	0.022772	-42.23	<2e-16	***
X7	-1.268362	0.023414	-54.17	<2e-16	***
X8	-0.804226	0.023657	-34.00	<2e-16	***
X9	-1.243891	0.022631	-54.96	<2e-16	***
X10	-1.951656	0.022676	-86.07	<2e-16	***
X11	0.408486	0.033222	12.30	<2e-16	***
X12	-3.184312	0.034787	-91.54	<2e-16	***
X13	-3.893966	0.085242	-45.68	<2e-16	***
X14	2.296857	0.116300	19.75	<2e-16	***
X15	-6.147652	0.142411	-43.17	<2e-16	***

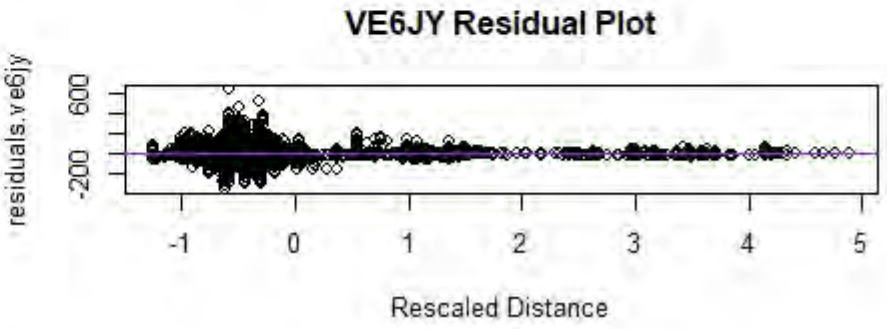
Final Model and Results

- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}{}^2 + \beta_5\, g(d(s))$



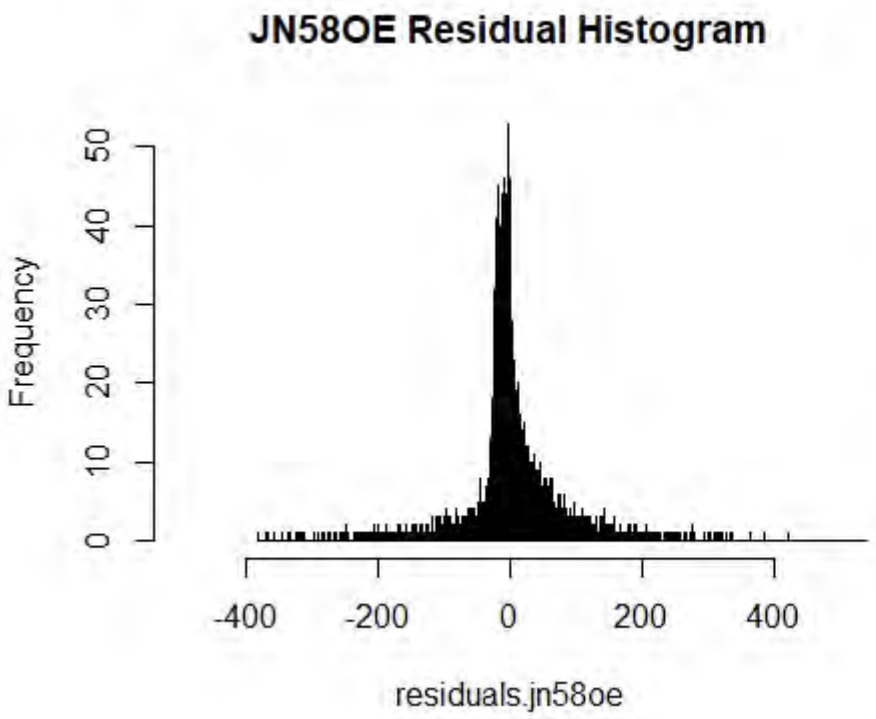


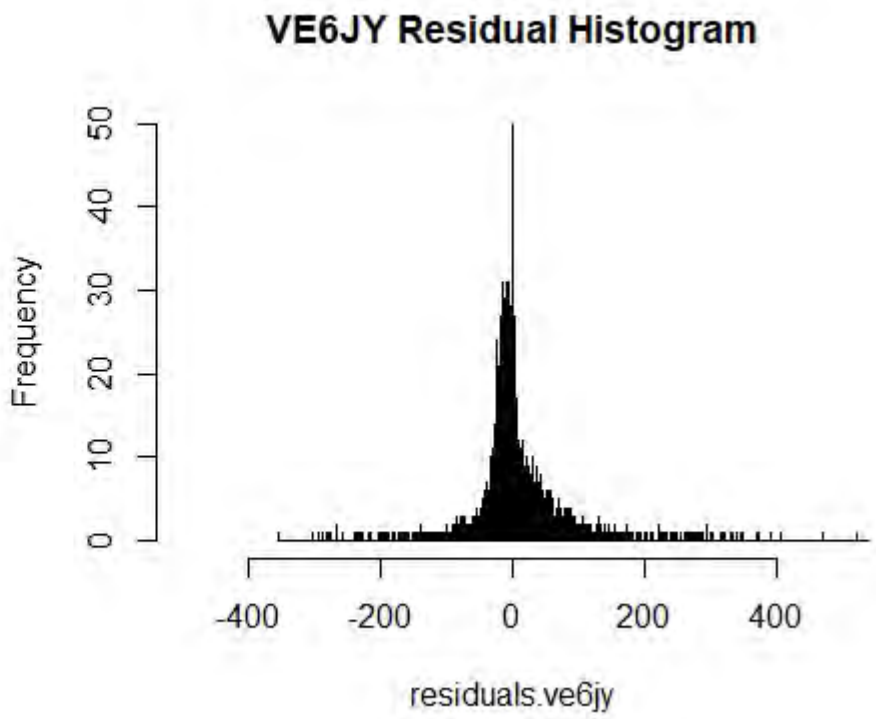




Final Model and Results

- Final Model: $\log(\lambda(t,s)) = \beta_1 X_{1,t} + \beta_2 X_{2,s} + \beta_3 X_{3,t,s} + \beta_4 X_{3,t,s}{}^2 + \beta_5\, g(d(s))$





CLEMSON UNIVERSITY

Conclusion and Future Work

- Additions to our statistical model
  - Expansion (frequencies, time, etc.)
  - Predictive checks
  - Bayesian methods
- Big data workflow
- Statistician + scientist teams
  - David Campbell: dtcampb@clemson.edu
  - Deborah Kunkel: dekunke@clemson.edu

CLEMSON UNIVERSITY

References

- Data sources:
  - Reverse Beacon Network. http://www.reversebeacon.net/index.php
  - Weak Signal Propagation Reporter Network. http://wsprnet.org/drupal/downloads
- Maidenhead coordinates lookup:
  - HA8TKS. https://dxcluster.ha8tks.hu/hamgeocoding/
- Map images:
  - Google Maps and affiliates. See image attributions.