Supplementary Information: Large-scale models based on population structure for the spatiotemporal distribution of U.S. porcine epidemic diarrhoea outbreaks

Eamon O'Dea Harry Snelson Shweta Bansal

July 26, 2015

Contents

Supplementary Figures

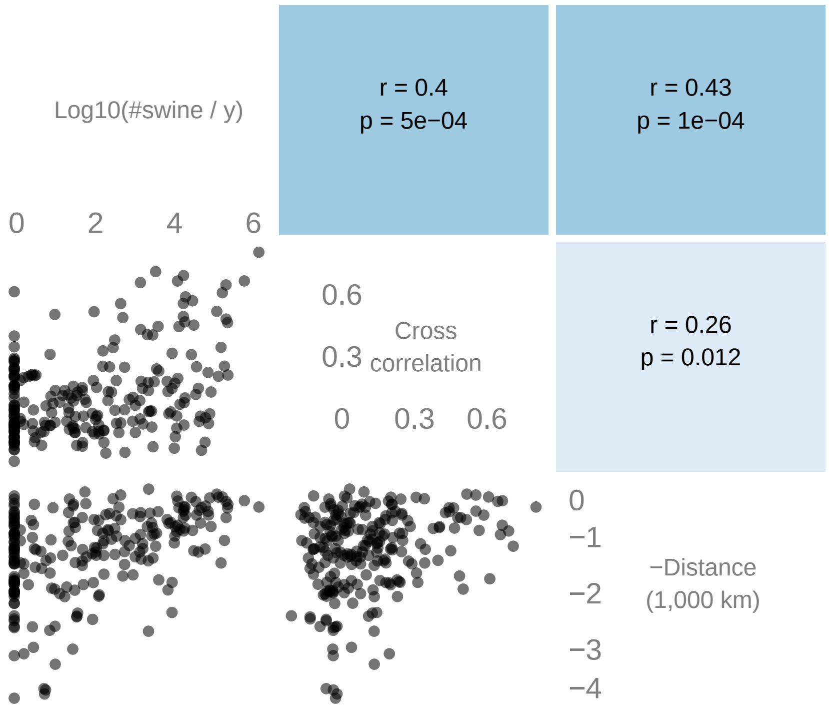


Figure S1: Scatter plots and Pearson correlations between pair-averaged (i.e., undirected) transport flows, cross correlations between time series of positive accessions, and negative geographic distances. The p values are from a Mantel tests.

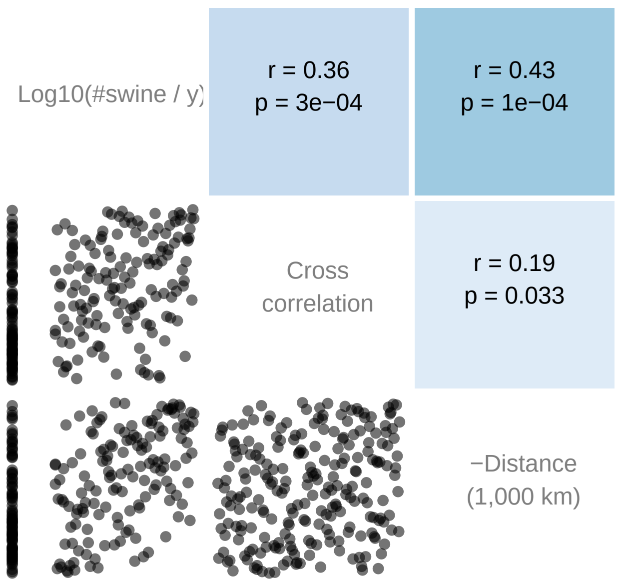


Figure S2: Rank scatter plots and Spearman correlations between transport flows, cross correlations between time series of positive accessions, and negative geographic distances. The p values are from a Mantel test.

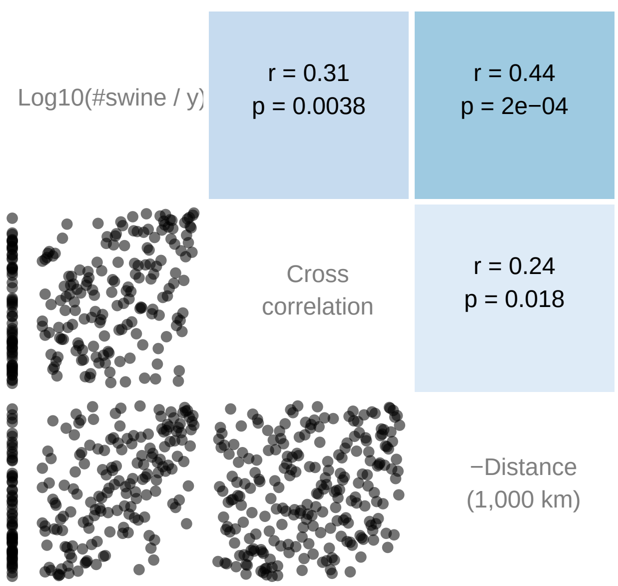


Figure S3: Rank scatter plots and Spearman correlations between pair-averaged (i.e., undirected) transport flows, cross correlations between time series of positive accessions, and negative geographic distances. The p values are from a Mantel test.

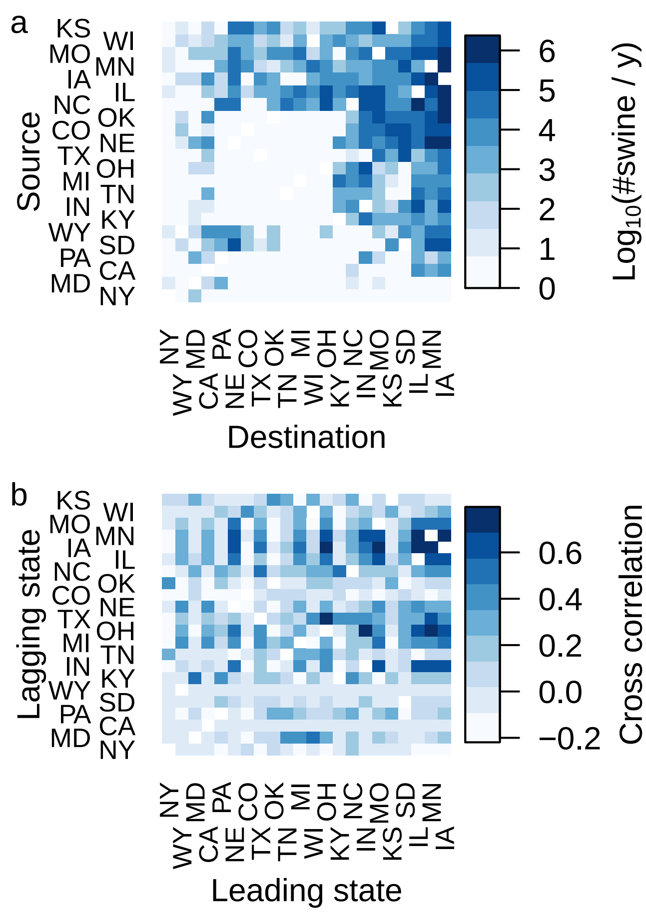


Figure S4: Comparison of swine transport flows and coupling of outbreak dynamics. (A) Annual swine transport flows from source states to destination states. (B) Cross correlations of PEDV positive accessions per week. Cross correlations are calculated as the correlations between positive accessions in the leading state with those in the lagging state in the previous week. Within-state values of flows and cross correlations are not included in the analysis and appear as white squares. In both panels, rows and columns are arranged to cluster together states with similar shipment flows.

Supplementary Table

Table S1: Farm types and the age classes of swine typically present on them. Ones (zeros) indicate the presence (absence) of an age class on a particular farm type.

Farms type	Suckling	Nursery	Grower/Feeder	Sow/Boar
Farrow to wean	1	0	0	1
Farrow to finish	1	1	1	1
Finish only	0	0	1	0
Farrow to feeder	1	1	0	1
Nursery	0	1	0	0

Supplementary Notes

1. Descriptive statistics of data sets

The tables below provide the number of observations n, the number of missing values, the number of unique values, and the mean. Depending on their distributions, variables are further described with some subset of quantiles, order statistics, histograms, and probability mass functions.

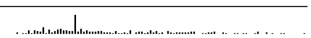
Variables in Mantel tests
3 Variables 231 Observations

Transport flow, log10(#swine / y + 1)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
231	0	113	1.644	0.0000	0.0000	0.0000	0.4771	3.4312	4.4564	5.0227

```
lowest : 0.0000 0.3010 0.4771 0.6021 0.6990
highest: 5.3076 5.3199 5.3958 5.5006 5.9350
```

Cross correlation in positive accessions



```
      n  missing   unique     Mean      .05      .10      .25      .50
    231        0      229   0.1189 -0.14330 -0.11563 -0.04456  0.03669
    .75      .90      .95
0.26464  0.45715  0.56794
lowest : -0.2078 -0.1994 -0.1791 -0.1762 -0.1670
highest:  0.6625  0.6840  0.7057  0.7188  0.7959
```

-Geographic distance (km)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
231	0	231	-1269	-2598.3	-2184.0	-1712.2	-1130.8	-723.3	-487.3	-367.1

```
lowest : -3834.1 -3760.3 -3690.0 -3667.8 -3242.6
highest:  -312.5  -296.8  -280.6  -243.1  -193.6
```

Variables used in stability selection with responses of litter rate and any positive accessions
26 Variables 42 Observations

Percent decrease in litter rate above 2

n	missing	unique
42	0	2

```
FALSE (33, 79%), TRUE (9, 21%)
```

Any positive accessions

n	missing	unique
42	0	2

```
FALSE (20, 48%), TRUE (22, 52%)
```

Log(#farms) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	40	5.474	2.962	3.556	4.580	5.277	6.757	7.548	7.609

```
lowest : 2.303 2.565 2.944 3.296 3.532
highest: 7.554 7.598 7.610 8.100 8.929
```

Log(mean over counties of #farms / km^2) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	-6.292	-9.164	-7.885	-7.248	-6.234	-4.987	-4.298	-3.873

```
lowest : -9.906 -9.572 -9.214 -8.218 -7.907
highest: -4.288 -4.282 -3.852 -3.774 -2.982
```

Log(mean over counties with farms of #farms / km^2) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	-5.915	-8.321	-7.534	-6.764	-5.954	-4.839	-4.203	-3.833

```
lowest : -8.586 -8.489 -8.349 -7.790 -7.571
highest: -4.199 -4.105 -3.819 -3.715 -2.982
```

Log(median over counties of #farms / km^2) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	40	-6.974	-10.795	-8.626	-8.016	-6.993	-5.699	-4.850	-4.369

```
lowest : -10.850  -9.752  -8.631  -8.588  -8.574
highest:  -4.814  -4.416  -4.366  -4.219  -3.115
```

Log(maximum over counties of #farms / km^2) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	-4.365	-7.400	-6.210	-5.301	-4.582	-3.299	-2.262	-1.834

```
lowest : -7.732 -7.506 -7.455 -6.364 -6.264
highest: -2.262 -2.058 -1.822 -1.563 -1.456
```

Log(swine inventory in year 2012) 

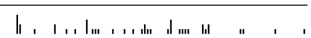
n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	5.327	1.496	1.807	3.790	5.301	7.077	8.219	8.917

```
lowest : 0.2624 0.9933 1.4951 1.5041 1.7918
highest: 8.2428 8.4338 8.9425 9.1050 9.9330
```

Log(pig crop) 

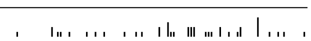
n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	6.007	1.731	2.281	4.274	6.573	8.038	8.953	9.408

```
lowest : 0.3365 0.5306 1.7192 1.9459 2.2721
highest: 8.9564 9.2229 9.4176 9.8014 9.9241
```

Log(inshipments) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	37	4.053	0.004766	0.182322	2.165058	4.429042	5.875295	6.714020	7.895151

```
lowest :  0.00000  0.09531  0.18232  0.69315  1.38629
highest:  6.71659  7.83320  7.89841  8.98457 10.08581
```

Log(marketings) 

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	6.224	1.866	2.184	4.802	6.612	8.174	9.001	9.699

```
lowest :  0.5878  1.8310  1.8563  2.0541  2.1518
highest:  9.0014  9.4175  9.7133  9.8706 10.6246
```

Weighted mean over counties of whether a county is in resource region 1

n	missing	unique	Mean
42	0	8	0.1672

```
0 (33, 79%), 0.334405144694534 (1, 2%), 0.661064425770308 (1, 2%)
0.667582417582418 (1, 2%), 0.705559368565546 (1, 2%)
0.796841785605831 (1, 2%), 0.857545839210155 (1, 2%), 1 (3, 7%)
```

Weighted mean over counties of whether a county is in resource region 2

n	missing	unique	Mean
42	0	6	0.1609

```
0 (33, 79%), 0.190403887033101 (1, 2%), 0.198352779684283 (1, 2%)
0.481981981981982 (1, 2%), 0.888992537313433 (1, 2%), 1 (5, 12%)
```

Weighted mean over counties of whether a county is in resource region 3

n	missing	unique	Mean
42	0	8	0.06456

```
0 (35, 83%), 0.0127543273610689 (1, 2%), 0.0453781512605042 (1, 2%)
0.332417582417582 (1, 2%), 0.37888198757764 (1, 2%)
0.425925925925926 (1, 2%), 0.516129032258065 (1, 2%), 1 (1, 2%)
```

Weighted mean over counties of whether a county is in resource region 4

n	missing	unique	Mean
42	0	7	0.08144

```
0 (36, 86%), 0.293557422969188 (1, 2%), 0.397515527950311 (1, 2%)
0.4 (1, 2%), 0.662087912087912 (1, 2%), 0.667447306791569 (1, 2%)
1 (1, 2%)
```

Weighted mean over counties of whether a county is in resource region 5

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	14	0.1038	0.00000	0.00000	0.00000	0.00000	0.03363	0.46252	0.65861

```
0 (29, 69%), 0.00576217915138816 (1, 2%), 0.018018018018018 (1, 2%)
0.0388349514563107 (1, 2%), 0.0960878517501716 (1, 2%)
0.111007462686567 (1, 2%), 0.115537848605578 (1, 2%)
0.142454160789845 (1, 2%), 0.337912087912088 (1, 2%)
0.476363636363636 (1, 2%), 0.525925925925926 (1, 2%)
0.665594855305466 (1, 2%), 0.825726141078838 (1, 2%), 1 (1, 2%)
```

Weighted mean over counties of whether a county is in resource region 6

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	11	0.123	0.0000	0.0000	0.0000	0.0000	0.0000	0.4974	0.8668

```
0 (32, 76%), 0.016597510373444 (1, 2%), 0.175257731958763 (1, 2%)
0.192037470725995 (1, 2%), 0.206611570247934 (1, 2%)
0.474074074074074 (1, 2%), 0.5 (1, 2%), 0.854368932038835 (1, 2%)
0.867469879518072 (1, 2%), 0.884462151394422 (1, 2%)
0.994237820848612 (1, 2%)
```

Weighted mean over counties of whether a county is in resource region 7

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	10	0.1353	0.0000	0.0000	0.0000	0.0000	0.0000	0.8352	0.8493

```
0 (33, 79%), 0.106796116504854 (1, 2%), 0.132530120481928 (1, 2%)
0.140515222482436 (1, 2%), 0.806451612903226 (1, 2%)
0.838383838383838 (1, 2%), 0.845238095238095 (1, 2%)
0.849462365591398 (1, 2%), 0.962962962962963 (1, 2%), 1 (1, 2%)
```

Weighted mean over counties of whether a county is in resource region 8

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	11	0.109	0.00000	0.00000	0.00000	0.00000	0.02778	0.45784	0.59870

```
0 (31, 74%), 0.037037037037037 (1, 2%), 0.150537634408602 (1, 2%)
0.154761904761905 (1, 2%), 0.161616161616162 (1, 2%)
0.193548387096774 (1, 2%), 0.22360248447205 (1, 2%)
0.483870967741935 (1, 2%), 0.574074074074074 (1, 2%), 0.6 (1, 2%)
1 (2, 5%)
```

Weighted mean over counties of whether a county is in resource region 9

n	missing	unique	Mean
42	0	5	0.05475

```
0 (38, 90%), 0.157676348547718 (1, 2%), 0.523636363636364 (1, 2%)
0.793388429752066 (1, 2%), 0.824742268041237 (1, 2%)
```

Log (positive accessions), weighted by shared border

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	34	3.591	-1.0000	-0.5756	0.3219	4.6585	6.2385	7.0276	7.3729

```
lowest : -1.0000 -0.5850 -0.4912 -0.2630  0.3219
highest:  7.0470  7.3038  7.3765  7.5107  7.6884
```

Log (positive accessions), weighted by directed flows

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	21.3	3.746	6.644	17.326	24.388	27.330	29.821	29.971

```
lowest :  2.459  3.047  3.644  5.687  6.384
highest: 29.823 29.924 29.974 30.071 30.639
```

Log (positive accessions), weighted by directed flows$^{0.5}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	14.74	2.594	3.714	12.226	17.094	18.529	19.861	19.913

```
lowest :  1.452  1.987  2.576  2.934  3.265
highest: 19.867 19.909 19.913 19.943 20.238
```

Log (positive accessions), weighted by directed flows$^{0.25}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	11.61	1.673	2.805	10.990	13.652	14.370	15.025	15.087

```
lowest :  0.9966  1.2395  1.6655  1.8149  2.4237
highest: 15.0411 15.0667 15.0884 15.1906 15.3049
```

Log (positive accessions), weighted by directed flows$^{0.125}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	10.13	1.097	2.674	9.857	11.869	12.503	12.716	12.774

```
lowest :  0.7848  0.8986  1.0941  1.1618  2.3620
highest: 12.7186 12.7648 12.7742 12.9473 13.0080
```

Log (positive accessions), weighted by directed flows$^{0.0625}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
42	0	42	9.428	0.8323	2.6110	9.4022	11.0280	11.5403	11.6722	11.7493

```
lowest :  0.6833  0.7381  0.8307  0.8624  2.3341
highest: 11.6724 11.7110 11.7513 11.8737 11.9198
```

Data for stability selection with response of total positive accessions
25 Variables 22 Observations

Log(total positive accessions)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	19	2.857	0.00000	0.06931	1.44208	3.05462	4.21639	5.46536	5.71144

```
-1.09861228866811 (1, 5%), 0 (2, 9%), 0.693147180559945 (1, 5%)
1.38629436111989 (2, 9%), 1.6094379124341 (2, 9%)
1.79175946922805 (1, 5%), 2.19722457733622 (1, 5%)
2.89037175789616 (1, 5%), 3.2188758248682 (1, 5%)
3.3322045101752 (1, 5%), 3.40119738166216 (1, 5%)
4.06044301054642 (1, 5%), 4.07753744390572 (1, 5%)
4.26267987704132 (1, 5%), 4.97673374242057 (1, 5%)
5.32787616878958 (1, 5%), 5.48063892334199 (1, 5%)
5.72358510195238 (1, 5%), 6.5206211275587 (1, 5%)
```

Log(#farms)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	6.481	4.044	5.096	5.648	6.701	7.436	7.609	8.075

```
lowest : 3.769 3.989 5.081 5.226 5.557
highest: 7.554 7.598 7.610 8.100 8.929
```

Log(mean over counties of #farms / km^2)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	-5.376	-7.455	-7.384	-6.036	-5.020	-4.400	-3.895	-3.778

```
lowest : -8.218 -7.458 -7.404 -7.205 -6.106
highest: -4.288 -4.282 -3.852 -3.774 -2.982
```

Log(mean over counties with farms of #farms / km^2)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	-5.168	-6.948	-6.809	-5.797	-4.860	-4.259	-3.847	-3.720

```
lowest : -7.790 -6.955 -6.811 -6.792 -5.967
highest: -4.199 -4.105 -3.819 -3.715 -2.982
```

Log(median over counties of #farms / km^2)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	-5.891	-8.323	-7.905	-6.317	-5.729	-5.182	-4.371	-4.226

```
lowest : -8.588 -8.344 -7.927 -7.708 -6.695
highest: -4.814 -4.416 -4.366 -4.219 -3.115
```

Log(maximum over counties of #farms / km^2)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	-3.474	-5.609	-5.366	-4.489	-3.301	-2.269	-1.845	-1.576

```
lowest : -6.364 -5.622 -5.371 -5.321 -4.846
highest: -2.262 -2.058 -1.822 -1.563 -1.456
```

Log(swine inventory in year 2012)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	6.825	4.203	4.474	5.760	7.064	7.985	8.892	9.097

```
lowest : 2.367 4.190 4.454 4.654 5.011
highest: 8.243 8.434 8.942 9.105 9.933
```

Log(pig crop)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	7.566	4.190	4.884	6.655	7.996	8.921	9.398	9.782

```
lowest : 3.138 4.159 4.787 5.753 6.510
highest: 8.956 9.223 9.418 9.801 9.924
```

Log(inshipments)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	5.711	2.492	2.779	4.518	5.825	6.681	7.892	8.930

```
lowest :  1.386  2.485  2.637  4.060  4.369
highest:  6.717  7.833  7.898  8.985 10.086
```

Log(marketings)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	7.781	5.021	5.548	6.672	8.102	8.987	9.684	9.863

```
lowest :  3.638  4.995  5.509  5.892  6.586
highest:  9.001  9.418  9.713  9.871 10.625
```

Weighted mean over counties of whether a county is in resource region 1

n	missing	unique	Mean
22	0	8	0.3192

```
0 (13, 59%), 0.334405144694534 (1, 5%), 0.661064425770308 (1, 5%)
0.667582417582418 (1, 5%), 0.705559368565546 (1, 5%)
0.796841785605831 (1, 5%), 0.857545839210155 (1, 5%), 1 (3, 14%)
```

Weighted mean over counties of whether a county is in resource region 2

n	missing	unique	Mean
22	0	6	0.2164

```
0 (15, 68%), 0.190403887033101 (1, 5%), 0.198352779684283 (1, 5%)
0.481981981981982 (1, 5%), 0.888992537313433 (1, 5%), 1 (3, 14%)
```

Weighted mean over counties of whether a county is in resource region 3

n	missing	unique	Mean
22	0	6	0.05433

```
0 (17, 77%), 0.0127543273610689 (1, 5%), 0.0453781512605042 (1, 5%)
0.332417582417582 (1, 5%), 0.37888198757764 (1, 5%)
0.425925925925926 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 4

n	missing	unique	Mean
22	0	6	0.1373

```
0 (17, 77%), 0.293557422969188 (1, 5%), 0.397515527950311 (1, 5%)
0.662087912087912 (1, 5%), 0.667447306791569 (1, 5%), 1 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 5

n	missing	unique	Mean
22	0	9	0.08424

```
0 (14, 64%), 0.00576217915138816 (1, 5%), 0.018018018018018 (1, 5%)
0.0960878517501716 (1, 5%), 0.111007462686567 (1, 5%)
0.142454160789845 (1, 5%), 0.337912087912088 (1, 5%)
0.476363636363636 (1, 5%), 0.665594855305466 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 6

n	missing	unique	Mean
22	0	4	0.07665

```
0 (19, 86%), 0.192037470725995 (1, 5%), 0.5 (1, 5%)
0.994237820848612 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 7

n	missing	unique	Mean
22	0	3	0.045

```
0 (20, 91%), 0.140515222482436 (1, 5%), 0.849462365591398 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 8

n	missing	unique	Mean
22	0	4	0.0431

```
0 (19, 86%), 0.150537634408602 (1, 5%), 0.22360248447205 (1, 5%)
0.574074074074074 (1, 5%)
```

Weighted mean over counties of whether a county is in resource region 9

n	missing	unique	Mean
22	0	2	0.0238

```
0 (21, 95%), 0.523636363636364 (1, 5%)
```

Log (positive accessions), weighted by shared border

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	4.962	0.417	2.229	3.898	5.531	6.719	7.369	7.504

```
lowest : -1.0000  0.3219  2.2224  2.2870  2.8074
highest:  7.0470  7.3038  7.3765  7.5107  7.6884
```

Log (positive accessions), weighted by directed flows

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	25.39	12.58	20.85	24.27	26.58	29.40	29.97	30.07

```
lowest :  8.984 12.157 20.718 22.022 24.010
highest: 29.823 29.924 29.974 30.071 30.639
```

Log (positive accessions), weighted by directed flows$^{0.5}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	17.46	8.835	15.139	17.202	18.190	19.646	19.913	19.942

```
lowest :  8.325  8.508 15.047 15.959 17.070
highest: 19.867 19.909 19.913 19.943 20.238
```

Log (positive accessions), weighted by directed flows$^{0.25}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	13.69	8.487	12.519	13.787	14.225	14.852	15.086	15.185

```
lowest :  7.058  8.277 12.461 13.034 13.732
highest: 15.041 15.067 15.088 15.191 15.305
```

Log (positive accessions), weighted by directed flows$^{0.125}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	11.91	8.416	11.263	12.110	12.307	12.641	12.773	12.939

```
lowest :  6.439  8.269 11.221 11.639 12.005
highest: 12.719 12.765 12.774 12.947 13.008
```

Log (positive accessions), weighted by directed flows$^{0.0625}$

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
22	0	22	11.05	8.384	10.649	11.253	11.390	11.641	11.707	11.866

```
lowest :  6.158  8.266 10.614 10.969 11.087
highest: 11.670 11.672 11.711 11.874 11.920
```

Distribution of positive accessions over states of origin by age class
4 Variables 20 Observations

Grower/Finisher

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	14	25.1	0.00	0.00	0.75	8.00	16.50	47.20	76.20

```
           0  1 2 4 6 10 11 12 13 27 41 45 67 251
Frequency  5  2 1 1 1  1  2  1  1  1  1  1  1   1
%         25 10 5 5 5  5 10  5  5  5  5  5  5   5
```

Nursery

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	15	19.95	0.00	0.00	0.75	6.50	15.50	52.20	75.85

```
           0  1 3 4 6 7 8 10 12 14 20 42 50 72 149
Frequency  5  2 1 1 1 1 1  1  1  1  1  1  1  1   1
%         25 10 5 5 5 5 5  5  5  5  5  5  5  5   5
```

Sow/Boar

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	12	9.7	0.00	0.00	0.00	2.50	8.25	37.20	39.60

```
           0 1 2 3 4  5 8 9 30 37 39 51
Frequency  8 1 1 1 1  2 1 1  1  1  1  1
%         40 5 5 5 5 10 5 5  5  5  5  5
```

Suckling

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	14	21.8	0.0	0.9	2.0	5.0	20.5	59.5	101.6

```
           0  1  2 3 4 6 10 11 20 22 52 55 100 132
Frequency  2  2  4 1 1 1  1  2  1  1  1  1   1   1
%         10 10 20 5 5 5  5 10  5  5  5  5   5   5
```

Variables in logistic regression of the proportion of positive accessions from the suckling age class
3 Variables 20 Observations

Observed proportion suckling

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	15	0.4045	0.0000	0.1385	0.2266	0.2679	0.4306	1.0000	1.0000

```
0 (2, 10%), 0.153846153846154 (1, 5%), 0.16 (1, 5%)
0.212765957446809 (1, 5%), 0.231173380035026 (1, 5%)
0.244444444444444 (1, 5%), 0.25 (1, 5%), 0.263157894736842 (2, 10%)
0.272727272727273 (1, 5%), 0.373134328358209 (1, 5%)
0.379310344827586 (1, 5%), 0.37956204379562 (1, 5%)
0.407407407407407 (1, 5%), 0.5 (1, 5%), 1 (4, 20%)
```

Number of positive accessions with known age class

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	18	76.55	1.00	1.00	3.75	23.50	64.00	214.90	283.15

	1	2	3	4	6	8	13	22	25	29	38	45	54	94	137	209	268	571
Frequency	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
%	15	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Model sampling probability of suckling

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
20	0	20	0.1313	0.06501	0.08031	0.09393	0.11338	0.13888	0.20801	0.27182

```
lowest : 0.05961 0.06530 0.08198 0.08972 0.09151
highest: 0.14607 0.19812 0.20097 0.27139 0.28010
```

Variables used in likelihood ratio test of within-state flows

7 Variables 1776 Observations

Positive accessions this week

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
1776	0	36	1.153	0	0	0	0	0	2	6

```
lowest :  0  1  2  3  4, highest: 54 61 65 66 96
```

Log(positive accessions last week + 0.5)

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
1776	0	35	-0.3485	-0.6931	-0.6931	-0.6931	-0.6931	-0.6931	0.9163	1.8718

```
lowest : -0.6931  0.4055  0.9163  1.2528  1.5041
highest:  3.9416  4.1190  4.1821  4.1972  4.5695
```

Scaled log[(median farm density among counties having farms) (# farms)^2]

```
       n   missing    unique      Mean       .05       .10       .25
    1776         0        48 1.409e-17  -1.11739  -0.83959  -0.43886
     .50       .75       .90       .95
 0.03578   0.56114   0.90760   0.98326
lowest : -1.6059 -1.1671 -1.1174 -0.9422 -0.8396
highest:  0.9076  0.9698  0.9833  1.0571  1.4788
```

Scaled log[2 (#swine moved within state) / y]

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
1776	0	48	1.949e-16	-1.1805	-1.0482	-0.4992	0.1086	0.5008	0.8120	1.0686

```
lowest : -1.3168 -1.2602 -1.1805 -1.1172 -1.0482
highest:  0.8120  0.8454  1.0686  1.2861  1.3881
```

Centred week

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
1776	0	37	0.5	-16.5	-14.5	-8.5	0.5	9.5	15.5	17.5

```
lowest : -17.5 -16.5 -15.5 -14.5 -13.5
highest:  14.5  15.5  16.5  17.5  18.5
```

State

n	missing	unique
1776	0	48

```
lowest : AL AR AZ CA CO, highest: VT WA WI WV WY
```

Centred offset

```
       n   missing    unique      Mean       .05       .10       .25
    1776         0       337 9.502e-17  -1.64789  -1.51607  -0.92311
     .50       .75       .90       .95
-0.03373   0.92442   1.58652   2.10097
lowest : -2.502 -2.496 -2.486 -2.479 -2.472
highest:  1.587  1.745  2.101  2.371  2.643
```

---

2. Detailed data descriptions

Transport flows

Our data on shipment of live swine was an estimate of the total number of swine moved between all ordered pairs of states over the course of a year. We refer to these estimates as transport flows. They were generated by the USDA Economic Research Service [1] and are available on the web.[1]

State-to-state distances

R's[2] datasets package provided the coordinates of the approximate geographic centres of each state. We used these coordinates to calculate the distance between pairs of states via the haversine formula. This formula uses a spherical model of the Earth to account for the curvature of the Earth's surface on the shortest path along the surface between two locations.

Positive accessions and litter rates

We obtained the case data from the January 8 version of a publicly available report of laboratory testing activity [3] via the "Number of New Cases Reported" link on the webpage of the American Association of Swine Veterinarians (AASV). We used the data in Tables 2 and 4 in this report. Table 2 contained the number of positive diagnostic case submissions stratified by the state of origin and the week of submission. Table 4 contained the number of positive diagnostic case submissions stratified by the state of origin and the age class of the swine from which the tissue sample was taken.

The precise meaning of the values in the time series changed in the week of June 16. Prior to that week, the values are the number of farms testing positive for PEDV. Beginning with that week, the values are the number of diagnostic cases submissions, or accessions, and there may be many of these accessions for each infected farm. However, in the 9 weeks prior to June 16, the difference between numbers of positive case submissions and numbers of positive farms was typically less than 5. Thus it seems that the number of case submissions approximates the number of infected farms reasonably well. As mentioned in the main text, positive accessions are correlated with the number of positive farms available in more recent reports.

The litter estimates are generally considered accurate but have limitations as indicators of PEDV burdens. For example, the 2013–2014 winter was unusually harsh and may explain some fraction of the losses. Also, the survey was not designed to estimate PEDV losses by state and may not have obtained a representative sample for each state.

Predictors of cumulative burden

In the fitted regression models for cumulative burdens, many states in the Northeast had close centroids and very similar residuals, suggesting a lack of independence at this spatial scale. Using single linkage clustering, we found two groups of states that formed chains of states with centroids less than 175 km apart: (1) MD, DE, and NJ; and (2) VT, NH, CT, MA, and RI. Because of the lack of independence among data from these states suggested by our initial fits, we created a reduced data set where the values of both predictive variables and response variables for these two groups of states were averaged to form single observations. The results presented are based on that data, for which no spatial autocorrelation was indicated by maximum likelihood fits of a model of exponentially decaying covariance in the residuals.

Counts of farms of different sizes were obtained from a database application available from the USDA [4]. This application contains data from the 2007 Census of Agriculture.

The balance sheet variables for each state came from estimates for the period of December 2011 to December 2012 in Ref. [5]. These variables were swine inventory, which is the total number of swine; pig crop, which is the number of pigs born that survive the first few weeks of life; inshipments, which is the number of swine imported to the state; and marketings, which is the number of swine either exported from the state or slaughtered at a commercial facility.

Farm resource regions classify counties into one of 9 general groups based on a wide variety of criteria including farm characteristics and crops and livestock produced [6]. A list mapping counties to these regions

[1] `http://webarchives.cdlib.org/sw1rf5mh0k/http://ers.usda.gov/Data/InterstateLivestockMovements/StateShipments.xls`

was obtained from a USDA spreadsheet.[2] We included a variable for each region and each state was assigned a value of in [0,1] for each such variable that was equal to the proportion of farms with 25 or more swine in the counties of that region.

Nearby positive accessions were calculated in various ways based on different possible models of spread. To represent cumulative exposure from a spatial model, we calculated for each state the average number of mean weekly positive accessions in other states with shared borders and used the logarithm of this average as a potential predictor. To represent cumulative exposure via shipment of pigs, we calculated for each state a weighted sum of mean weekly positive accessions in other states, where the weights were given by the flows from that state. Because we found with our simulations that the dependence of positive accessions in one area can be expected to increase sublinearly with the flows from that area, we used various power transforms of the flows as weights. Specifically, we used the flows raised to the powers of 1, 1/2, 1/4, 1/8, and 1/16. These weighted sums were then log transformed to create a series of potential predictors.

The summary statistics for farm density were average number of farms per county, average number of farms in counties with at least one farm, median number of farms per county, and maximum number of farms per county. Some states had a median of zero farms per county. When log transforming, one half of the smallest positive median was added to values of all states before transformation. For this analysis, we defined farms as operations with 25 or more swine.

3. Age-specific reporting bias

Because infection mortality is high among piglets only [7], we might expect that operations without piglets are less likely to perform diagnostic testing. We can gain some insight into such potential reporting biases from the data about the age classes of diagnostic samples. These age classes are suckling (less than 1 month old or still on sow), nursery (1–3 months of age), grower/finisher (3–8 months), and sow/boar (more than 8 months old). Although the report providing the data uses the term age class and gives those particular age ranges for each class, these terms are really names for production stages in the swine industry for which there may be some variation outside of those ranges, in particular for the time at which pigs are weaned and sent to a nursery.

We tested the hypothesis that, among those states having any positive accessions with known age class, the age-class distribution is independent of the state of origin. Supplementary Note 1 contains descriptive statistics of the state distributions for each age class. We used simulation to generate a null distribution of test statistics rather than rely on asymptotic results because several of the observed cell counts were small. Tables of counts of samples in all combinations of age classes and states were simulated under the hypothesis of independence of age-class and state-of-origin. The simulated tables had the same marginal distributions as the observed data. The sum of Pearson residuals based on observed and expected cell counts was our test statistic. We conducted a test of independence at a significance level of 0.05. We rejected the hypothesis of independence based on an observed test statistic of 210, which was greater than the test statistic in all 10,000 of our null statistics ($p < 1 \times 10^{-4}$). To quantify the extent of dependence, we calculated that the uncertainty coefficient[8] of age class, given the state of origin, was equal to 0.05, which indicates relatively weak dependence.

To determine whether the proportion of positive accessions in the suckling age class may be explained by the distribution of farm types within a state, we compute expected proportions under a model of two-step random sampling as follows. In the first step, we sample a certain type of farming operation from a distribution of operation types. Table S1 gives the names of the available types. We obtain the distribution of these types for each state from census data[4]. In the second step we draw an age class from the age class distribution of the sampled farm type.

We derive an age-class distribution by first assuming that sows on average produced 2.31 litters of weaning size 10.3 every year and that pigs spent 21.5 days as suckling pigs, 46.0 days in the nursery stage, and 121.5 days in the grower/finisher stage. Those parameters are taken from 2012 averages from sow farms, nurseries, and conventional finishing farms participating in a U.S. benchmarking system [9, Tables 2, 4, and 5]. Larger farms tend to use artificial insemination [10, Table 3] and thus we assume that boars make up a negligible part of the total population on sow farms.

[2] `http://www.ers.usda.gov/Briefing/ARMS/resourceregions/reglink.xls`

Given these parameters, we calculate an age-class distribution for the entire population by first calculating the rate of weaned pig production from the number of sows. The number of animals in all of the other age classes follow as the product of that rate and the average time spent in each class. Age classes on a farms with some subset of age classes follow as a subset of the age-class distribution for the entire population to those classes present on the farm. Table S1 shows which age classes are typically present on each type of farm. We normalise these age-class distributions to obtain sampling probabilities conditional on a farm type.

We used a standard logistic regression analysis to test for an association between observed proportions of pigs in the suckling age class and those proportions predicted by our sampling model. The response variable was the whether or not positive accessions where in the suckling age class and the predictors were probabilities from our sampling model and an intercept. We conducted a two-tailed Wald test of the hypothesis that the regression coefficient for the sampling probabilities was zero, and failed to reject this hypothesis ($p = 0.64$). See Supplementary Note 1 for descriptive statistics of the variables in the regression.

To see if the expected and observed probabilities were different on average, we fitted an intercept-only logistic model with logits of expected probabilities as offset terms. The observed log odds of suckling positive accessions was on average 3.54 natural logarithmic units above those predicted by random sampling (95% profile confidence interval = [3.15,3.97]). Removing highly influential observations (i.e., IA, NC, OK, KS, IL, and MN) resulted in somewhat lower interval estimate bounds of [1.86, 3.51]. These results indicate that farms with unweaned pigs are either more likely to choose unweaned pigs to be diagnostic samples than other pigs, more likely to seek laboratory confirmation of PEDV, or more likely to experience an outbreak than other farms.

4. Transmission model details

A number of simplifying assumptions are implicit in our main equation

$$\mathrm{E}(I_{i,t+1}) = \beta_{i,t}(\textstyle\sum_j w_{i,j} I_{j,t} + \eta)^\alpha S_{i,t}.$$

First, we treat entire farms as either infective or susceptible. This is consistent with the observation that infection may spread rapidly following the first appearance of clinical signs [11, 12, 13, 14, 15]. The spread may also be actively promoted as a part of recommended feedback procedures to establish herd immunity [16]. Second, we assume that the infection of a farm lags 1 week behind its infectious exposure. Consistently, the time from the introduction of infected animals to the appearance of clinical symptoms in PEDV-naive herds is typically less than 7 days [16, 13]. We also found that a 1-week lag had a higher likelihood in our models than lags of 2 to 4 weeks. Third, we assume that farms are only infectious for 1 week. On the one hand, virus shedding from individuals has been observed in experimental settings to subside within 9 days of infection [17], and an infected animal's diarrhoea has been observed in the field to typically last for 5 days [11]. On the other hand, it can take affected farms several weeks to return to baseline production levels [11, 13]. Our assumption is that herds will be most infectious the first week, perhaps because the number of animals shedding later becomes smaller or because more stringent biosecurity reduces the amount of infectious material leaving the farm. This assumption is congruent with those made by Ref. [18, p. 71] in setting parameters for an agent-based model of spread.

For the number of farms N_i, we used data from the 2002 Census [19] instead of data from more recent censuses so as to obtain farm count data that were contemporary with the transport flow data, which are from 2001 [1]. In this analysis, we included farms with any swine in the counts, unlike our analysis of cumulative burdens where we only included farms with at least 25 swine. All farms were included here because farms with fewer than 25 swine are numerous enough to constitute a non-negligible fraction of total swine inventory and flows.

References

[1] USDA ERS. Interstate livestock movements. By D. Shields and K. Mathews. Available: `http://www.ers.usda.gov/publications/ldpm-livestock,-dairy,-and-poultry-outlook/lpdm10801.aspx#.U26fN1Qt5Mk` (2003). Accessed 14 November 2013.

[2] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014).

[3] USDA APHIS VS. Porcine epidemic diarrhea virus (PEDv) testing data from NAHLN laboratories (2014). Available: http://www.aasv.org/pedv/PEDV_weekly_report_140108.pdf. Accessed 14 January 2014.

[4] USDA NASS. 2007 Desktop Data Query Tool 1.02 (2009). Available: http://www.agcensus.usda.gov/Publications/2007/Online_Highlights/Desktop_Application/dataquery.zip. Accessed 14 January 2013.

[5] USDA NASS. Meat Animals Production, Disposition, and Income 2012 Summary (2013). Available: http://usda01.library.cornell.edu/usda/current/MeatAnimPr/MeatAnimPr-04-25-2013.zip. Accessed 31 July 2013.

[6] USDA ERS. Farm Resource Regions. Available: http://www.ers.usda.gov/ersDownloadHandler.ashx?file=/media/926929/aib-760_002.pdf (2000). Accessed 3 April 2014.

[7] USDA. Summary of PEDV Actions. Available: http://www.usda.gov/documents/pedv-summary-actions.pdf (2014). Accessed 10 May 2014.

[8] Tang, W., He, H. & Tu, X. M. *Applied Categorical and Count Data Analysis* (CRC Press, 2012).

[9] Kenneth J. Stalder. Pork industry productivity analysis. Available: http://old.pork.org/filelibrary/research/ipafull.pdf (2013). Accessed 2 December 2014.

[10] USDA ERS. The changing economics of U.S. hog production. By N. Key and W. McBride (2007). ERR-52. Available: http://www.ers.usda.gov/media/244843/err52.pdf. Accessed 2 December 2014.

[11] Martelli, P. *et al.* Epidemic of diarrhoea caused by porcine epidemic diarrhoea virus in Italy. *Vet Rec* **162**, 307–310 (2008).

[12] Stevenson, G. W. *et al.* Emergence of porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J. Vet. Diagn. Invest.* **25**, 649–654 (2013).

[13] EFSA AHAW Panel. Scientific opinion on porcine epidemic diarrhoea and emerging pig deltacoronavirus. *EFSA Journal* **12**, 3877 (2014).

[14] Bowman, A. S., Krogwold, R. A., Price, T., Davis, M. & Moeller, S. J. Investigating the introduction of porcine epidemic diarrhea virus into an Ohio swine operation. *BMC Vet Res* **11**, 38 (2015).

[15] Jung, K. & Saif, L. J. Porcine epidemic diarrhea virus infection: etiology, epidemiology, pathogenesis and immunoprophylaxis. *Vet J* **204**, 134–143 (2015).

[16] Geiger, J. O. & Connor, J. F. Porcine epidemic diarrhea, diagnosis, and elimination. Available: https://www.aasv.org/aasv%20website/Resources/Diseases/PED/13-05-29PEDWhitePaper.pdf (2013). Accessed 23 January 2015.

[17] de Arriba, M. L., Carvajal, A., Pozo, J. & Rubio, P. Mucosal and systemic isotype-specific antibody responses and protection in conventional pigs exposed to virulent or attenuated porcine epidemic diarrhoea virus. *Vet Immunol Immunopathol* **85**, 85–97 (2002).

[18] ANSES. Relatif au risque dmergence de la diarrhe pidmique porcine (DEP) due un nouveau variant du virus de la DEP en France. Available: http://www.lafranceagricole.fr/var/gfa/storage/fichiers-pdf/Docs/2014/DEP-28juillet.pdf. Accessed 12 January 2015.

[19] USDA NASS. 2002 Census of Agriculture Query Tool. Available: http://www.agcensus.usda.gov/Publications/2002/Download_Data_Query_Application/agcensus2002.zip. Accessed 23 December 2014.