

11-1-2015


Contrails: Causal Inference Using Propensity Scores

Dean S. Barron

twobluecats.com, dean@twobluecats.com

Joe H. Brown

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Barron, Dean S. and Brown, Joe H. (2015) "Contrails: Causal Inference Using Propensity Scores," *Journal of Modern Applied Statistical Methods*: Vol. 14 : Iss. 2 , Article 18.

DOI: [10.22237/jmasm/1446351420](https://doi.org/10.22237/jmasm/1446351420)

Available at: <http://digitalcommons.wayne.edu/jmasm/vol14/iss2/18>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Contrails: Causal Inference Using Propensity Scores

Cover Page Footnote

Author contributions The seminal concept of subjecting contrail related data to causal inference was by J.H.B., and subsequently developed and discussed by J.H.B. and D.S.B. The logistic regression was coded in SAS by J.H.B. The remainder of the research, SAS coding, and all writing was performed by D.S.B.

Contrails: Causal Interference Using Propensity Scores

Dean S. Barron
twobluecats.com
Long Beach, CA

Joe H. Brown
Irvine, CA

Contrails are clouds caused by airplane exhausts, which geologists contend decrease daily temperature ranges on Earth. Following the 2001 World Trade Center attack, cancelled domestic flights triggered the first absence of contrails in decades. Resultant exceptional data capacitated causal inference analysis by propensity score matching. Estimated contrail effect was 6.8981°F.

Keywords: Contrails, contrails effect, airplane exhaust, causal inference, propensity score, resampling, logistic regression, regression, MCMC

Introduction

Contrails are the clouds formed as a result of the introduction of relatively warm water vapor from airplane engine exhausts into surrounding cold, moist, atmospheric air (the word "contrails" is a contraction of two words, "condensation trails.") Under salient conditions, such mixing within the airplane engine exhaust plume saturates the atmospheric air, causing condensation of water droplets upon the exhaust particles. In turn, these newly formed droplets freeze into ice particles that constitute contrails (Schumann, 2005; EPA, 2000). The process also depends on non-atmospheric factors, such as engine and fuel characteristics (Wendler & Stuefer, 2002). The contrail formation process typically occurs at altitudes over 25,000 ft. and temperatures below -40°C .

Geologists asserted that contrails (1) decrease the daily high temperature by blocking incoming sunlight, (2) increase nightly low temperatures by preventing escape of greenhouse gases, and, therefore (3) decrease the daily temperature range on the Earth's surface below (e.g., Meerkotter, et al., 1999). This contrails effect was estimated to be 1.98°F or 3.24°F (Travis, et al., 2002); the greater of these is hereafter referred to as the Travis estimate.

Dean Barron is a statistical consultant at twobluecats.com. Email him at dean@twobluecats.com.

A contrail may dissipate quickly or linger for hours. Persistent contrails may grow expansively and then frequently morph or incorporate into cloud cover (EPA, 2000). Over individual geographic areas, the presence of contrails depends on existent conditions.

Therefore, observing the actual temperature range in the absence of contrails was impossible in areas where contrails had always been present. After the World Trade Center attacks of 11 September 2001, however, all flights in the United States were suspended for several days. Thus, a complete absence of contrails prevailed, including those locations where contrails had been present continuously for decades.



Figure 1. Contrail (Barron, 2013)

CONTRAILS: CAUSAL INTERFERENCE

Purpose of the Study

Data situations with such counterfactuals are precisely the forte of analysis using causal inference. A propensity score (PS) was modeled and then used to match from the control group without replacement for the treatment group. Additionally, regression analysis and Bayesian Markoff Chain Monte Carlo (MCMC) were performed.

Data were obtained from The National Climatic Data Center (NCDC), which had daily historical data since 1929 from approximately 300 countries and 30,000 cities. The treatment group was defined as United States (hereafter, referred to as, "domestic") stations data from September 12-13, 2001, taking advantage of the absence of contrails. The control group was defined as all non-treatment station readings, both domestic and international. The data was subjected to random sampling and quality control.

The contrails effect, which, in causal inference terminology is the Average Treatment Effect on the Treated (ATT), was estimated to be 6.8981°F ($p < 0.0001$), compared with 6.5513°F ($p < 0.0001$) from the naive regression, and 6.5195°F ($\alpha = 0.05$ HPD Interval 5.7795, 7.2552) from MCMC simulation. All were more than twice the Travis estimate. The propensity score matching approach was determined to be preferable due to its superior covariate characteristics.

Methodology

Data

The NCDC weather-related database stores daily data as collected by the National Weather Service (NWS) Automated Surface Observing System (ASOS) in downloadable .txt format inside triple-compressed op-op.gz-tar formatted files (NCDC, 2010). The study data were restricted to measurements from stations that were operational in 2001.

These observations were further limited to 0-4 weeks before and after each September 12-13 for each of the three superimposable calendar years 1990, 2001, and 2007. Treatment variable, *CONTRAILS0*, was defined:

$$\begin{aligned} \text{CONTRAILS0} = \begin{cases} 1, & \text{treatment, domestic 12 SEP2001 and 13 SEP2001} \\ 0, & \text{control, otherwise} \end{cases} \quad (1) \end{aligned}$$

Contrail formations above airports have different characteristics than above non-airport locations. Because contrails generally do not form until aircraft reach 25,000 foot altitudes, contrails above airports typically derive from aircraft flights which had originated from other airports. Hence, airports might or might not have contrails (Mims, Chambers & Oostra, n.d.). Therefore, for this analysis, all airports were excluded from the control dataset only.

A two-stage stratified random sampling scheme was then imposed. Domestic data formed the first group. The United States was the only nation that stopped flights, therefore, neighboring Mexico and Canada formed the second group. Belgium and France were chosen as European counterparts for the third group. All other countries constituted the fourth group.

The first sampling stage selected 1,607 stations as treatment and 8,805 as control; from this, the second random sampling stage selected 278 and 440, respectively. The latter corresponded to a possible 3,214 and 478,250 observations, respectively. This data sampling procedure was designed to facilitate the required manual identification and subsequent elimination of airport locations.

Resultant samples sizes contained 556 treatment and 22,810 control observations, of which only 503 treatment and 4,737 control actually contained data. Further quality control on missing critical variables (*dewp*, *slp*, *wdsp*, *visib*, and *temperature-related*), dropped the final analysis dataset to 322 treatment and 2,557 control observations.

In addition to the variables contained in the NCDC database, the adjusted latitude was calculated using the formula (2) to correct for gravity (Bauer, et al., 2000). Normal gravity is defined as the gravity which would be observed were planet Earth to be a perfect ellipsoid with associated perfect rotation. The corrected latitude reflects deviations from ideal conditions, and is a function of only the latitude.

$$latitudecorr = 9.78 \times 10^5 * \left(\begin{array}{l} 1 + 5.28 \times 10^{-3} * \sin^2(latitude) \\ + 2.35 \times 10^{-5} * \sin^4(latitude) \end{array} \right) \quad (2)$$

Variables that were included in the propensity logistic regression model are described in Table 1.

CONTRAILS: CAUSAL INTERFERENCE

Table 1. Variables

ID	Variable	Required non- missing	Description
1	<i>CONTRAILS0</i>		1 = TREATMENT, Absence of contrails 0 = CONTROL, contrails present / contrails effect
2	<i>temp</i>	YES	mean temperature for the day in degrees Fahrenheit
3	<i>dewp</i>	YES	mean dew point for the day in degrees Fahrenheit
4	<i>slp</i>	YES	mean sea level pressure for the day
5	<i>visib</i>	YES	mean visibility for the day in miles
6	<i>wdsp</i>	YES	mean wind speed for the day in knots
7	<i>MXSPD</i>		maximum sustained wind speed
8	<i>PRCP</i>		total precipitation
9	<i>p133fog</i>		fog / FRSHTT character 1
10	<i>p134rain</i>		rain or drizzle / FRSHTT character 2
11	<i>p135snow</i>		snow or ice pellets / FRSHTT character 3
12	<i>p137thun</i>		thunder / FRSHTT character 5
13	<i>elev</i>		elevation in meters
14	<i>latitudecorr</i>		absolute value latitude in degrees
15	<i>latitudeabs</i>		latitude correction for gravity in milligalileos
16	<i>temprange</i>	YES	temperature range in degrees Fahrenheit

Analysis

Causal inference, regression analysis, and Bayesian MCMC were used. The several shades of each resulted in a total of 10 different methods, hereafter referred to as *METHOD1* through *METHOD10*.

Causal Inference

The Propensity Score (PS) was the predicted value from the linear first order logistic regression model of *CONTRAILS0* as a function of the covariates. All variables were retained to maximize R^2 .

For *METHOD1*, the PS of a treatment observation was compared with the PS of any remaining unmatched control observation. Matching by the absolute smallest PS difference, a greedy strategy was implemented in descending PS order of treatment observations. The ATT estimate for *CONTRAILS0* was equal to the *temprange* difference of treatment and control groups from the matched observation pairs, and evaluated by t test.

In *METHOD2*, resampling was performed to examine if the dataset perhaps had yielded a coincidentally favorable match. Nine treatment group sample sizes, $ntrt$, (288, 216, 162, 136, 108, 96, 81, 72, 68) were resampled ($n = 180$) at a corresponding specified control to treatment observational ratio (2, 3, 4, 5, 6, 7, 8, 9, 10, respectively). Because there was also potential for relative abundance of a control subregion to impact results, each of the four control subregions were equally represented, as calculated in (3).

$$ncntl_{\text{subregion}} = (1/4) * (ntrt) * (\text{control-to-treatment observational ratio}) \quad (3)$$

Care was taken to select whole numbers and ensure that $ncntl_{\text{subregion}} < 174$, because that was the sample size of the smallest subregion.

For each individual resample, the ATT was calculated identically as in *METHOD1*. For each $ntrt$ level, the ATT was calculated as the mean of its 180 samples; the overall ATT was the mean of the 1,620 runs.

In *METHOD3*, the tails of the dataset were trimmed to only the region of overlapping PS ranges of the treatment and control observations. The PS minima and maxima were determined for treatment ($PSmin_{trt}$, $PSmax_{trt}$) and control ($PSmin_{cntl}$, $PSmax_{cntl}$). A new PS range was set from the maximum minimum ($\max(PSmin_{trt}, PSmin_{cntl})$) to the minimum maximum ($\min(PSmax_{trt}, PSmax_{cntl})$) by dropping external values. In *METHOD4*, resampling was also performed.

For the best among the four methods, the resultant matched pairs and frequency distributions of the selected countries were analyzed. Patterns of the matched pairs were noted.

Regression

Three regressions were conducted to provide baseline comparisons for the propensity matching results, and to provide parameter estimates for other variables (4, 5, 6).

$$temprange = f(\text{CONTRAILS0, full model with all variables}) \quad (4)$$

$$temprange = f(\text{CONTRAILS0, best stepwise/backward elimination result}) \quad (5)$$

$$temprange = f(\text{CONTRAILS0}) \quad (6)$$

CONTRAILS: CAUSAL INTERFERENCE

Hereafter these are referred to as *METHOD5*, *METHOD6*, and *METHOD7*, respectively. Resampling was performed on the best of the three, hereafter, referred to as *METHOD8*.

Bayesian

Two MCMC regression simulations were run, based upon (5) and (6), referred to as *METHOD9* and *METHOD10*, respectively. Blocking strategy was determined by a correlations and resultant convergence characteristics. Non-informative priors were implemented first. When not feasible, the parameter estimates from the corresponding regression were to be used as informative priors.

The *CONTRAILS* estimates from all methods and Travis were compared. The MCMC simulation *METHOD9* posterior estimates for *CONTRAILS0* were analyzed to determine the percentage that were greater than each *CONTRAILS0* estimate. The probability that a particular *CONTRAILS0* estimate was an underestimation corresponds to this percentage.

Covariate and contrail effect estimate comparisons

Covariate differences between the matched treatment and control groups were calculated to reveal differences between the groups, which were compared with differences from the analysis dataset. Transition from significant to not significant was used as evidence of amelioration of covariate mean differences.

Omnibus distribution tests

Distributional differences between treatment and control groups were subjected to omnibus tests. These were Kolmogoroff-Smirnoff (KS), Cramér-von Mises (CM), and “oando” (see the Master’s thesis of the first author, Barron, 2007).

Results

Causal Inference

Propensity Score Logistic regression for PS was performed including all covariates with intercept using the final dataset ($n_{ttl} = 2879$). The resultant model of *CONTRAILS0* was statistically significant ($X^2 = 289.0694$, $df = 14$, $p\text{-value} < 0.0001$). The area under the ROC curve $c\text{-value} = 0.785$, Somers' $D = 0.570$, Kendall's $Tau\text{-}a = 0.113$, and standard definition of percentage behavior explained by model, $R^2 = 0.1127$. All correlations with

CONTRAILS0 and maximum likelihood parameter estimates are detailed in Table 2.

Table 2. Correlations and Propensity Score (PS) Logistic Regression Results

ID	Variable	Correlation	Parameter Estimate	Wald χ^2	p-value
0	<i>Intercept</i>	N/A	-68.1469	3.7370	0.0532
1	<i>CONTRAILS0</i>	1	N/A	N/A	N/A
2	<i>temp</i>	0.1267	0.1246	109.8874	<0.0001
3	<i>dewp</i>	0.0225	-0.1089	89.6451	<0.0001
4	<i>slp</i>	0.0989	0.0801	36.5800	<0.0001
5	<i>visib</i>	-0.0942	-0.0795	35.8712	<0.0001
6	<i>wdsp</i>	-0.0581	-6.6053E-03	0.1791	0.6721
7	<i>MXSPD</i>	-0.0439	-3.4946E-03	2.4711	0.1160
8	<i>PRCP</i>	-0.0217	-0.1344	0.3487	0.5549
9	<i>p133fog</i>	0.0914	1.3716	49.4748	<0.0001
10	<i>p134rain</i>	-0.0449	0.2519	2.0147	0.1558
11	<i>p135snow</i>	-0.0404	-1.0575	1.0007	0.3171
12	<i>p137thun</i>	0.0871	1.3485	21.3320	<0.0001
13	<i>elev</i>	0.0063	-4.8630E-04	10.7487	0.0010
14	<i>latitudecorr</i>	0.0003	-1.6200E-05	0.2089	0.6476
15	<i>latitudeabs</i>	-0.0895	-2.0420E-02	4.5209	0.0335
16	<i>temprange</i>	0.3119	N/A	N/A	N/A

To determine if there would be sufficient PS coverage to enable matching of treatment and control, the PS range was divided into four bins with equal *n*-treatment counts. Spread was adequate (Table 3).

Table 3. Propensity Score (PS) Frequency Distributions by Bin

	PS RANGE	n _{trt}	n _{cntl}	RATIO
BIN1	0.0002, 0.1160	80	1790	22.38
BIN2	0.1160, 0.1724	81	357	4.41
BIN3	0.1724, 0.2725	81	267	3.30
BIN4	0.2726, 0.8380	80	143	1.79
TOTAL	0.0002, 0.8380	322	2557	7.94

CONTRAILS: CAUSAL INTERFERENCE

METHOD1 / Matched Pairs, No Resampling, No Overlap Mean difference of *temprange* between matched pairs, the ATT estimate, was 6.8981 ($t = 9.91$, $p < 0.0001$, 95%CI 5.5293, 8.2670). The mean absolute distance between matched propensity scores was 0.0035 (median < 0.0001 , range < 0.0001 , 0.1033).

METHOD2 / Matched Pairs, Resampling, No Overlap The results consistently approximate the ATT estimate obtained with the non sampled data. The 1620 runs from the 9 different combinations had *temprange* mean = 6.7871 (median = 6.7847, range 2.5779, 10.8118). The mean of PS matched mean absolute distances was 0.0194 (median = 0.0133, range 0.0005, 0.1040). The results of the runs of the *ntrt* and control to treatment observational ratios appear in Table 6.

METHOD3 / Matched Pairs, No Resampling, Overlap Trimming down to the overlap region reduced the dataset to $ntrt = 321$ $ncntl = 2525$ $nttl = 2846$. Compared with the analysis dataset, this was a reduction of only one treatment and 32 control observations. The mean difference of *temprange* was 6.8931 ($t = 9.88$, $p < 0.0001$), with a mean absolute distance between matched propensity scores of 0.0032 (median < 0.0001 , range < 0.0001 , 0.0978).

METHOD4 / Matched Pairs, Resampling, Overlap The contrail effect estimates were slightly higher than those without the overlap strategy. The 9 different combinations averaged *temprange* = 6.9654 (median = 6.9352, range 3.2071, 10.7119). The mean of PS matched mean absolute distances was 0.0141 (median = 0.0072, range 0.0005, 0.0960). The results of the runs of the various *ntrt* and control to treatment observational ratios are also summarized in Table 6.

Analysis of the matches

The majority of treatment - control pairs appeared either once or twice. There were 200 distinct ordered pairs within the 322 matches, of which 171 (85.50%) had fewer than three occurrences. Only four appeared five or more times, California-France (11), Texas-France (8), Texas-United States (6), and California-Mexico (5).

All four strata of control country groups were represented in the matches. Despite the boost in percent observations secondary to designation as separate subgroups, the Relative Risk (RR) of selection for CANADA/MEXICO and BELGIUM/FRANCE were only somewhat lower than OTHER

INTERNATIONAL. Not surprisingly, the UNITED STATES group had the highest RR, as in Table 4.

Table 4. Subgroup Counts in Control Data

STRATUM	COUNTRIES	<i>nmatch</i>	<i>ncntl</i>	row % <i>match</i>	RR	column % <i>match</i>	column % <i>cntl</i>
1	UNITED STATES	41	174	23.56%	2.00	12.73%	6.80%
2	CANADA/MEXICO	65	631	10.30%	0.77	20.19%	24.68%
3	BELGIUM/FRANCE	102	851	11.99%	0.93	31.68%	33.28%
4	OTHER INTL	114	901	12.65%	1.01	35.40%	35.24%
TOTAL		322	2557	12.59%			

Twenty-five countries were included in the control population. The highest percentage of matched control observations was 50% selected for Australia (*nmatch* = 9 *ncntl* = 18); the lowest was the lone 0% for Georgia (0/19). Over two-thirds had RR for selection between 1/3 and 3 (17/25).

Regression

METHOD5, METHOD6, AND METHOD7 / Naive Regression, No Resampling The full regression model (*METHOD5*) with all covariates was statistically significant ($R^2 = 0.5713$, $F = 254.32$, $p < 0.0001$) with a parameter estimate for *CONTRAILS0* of 6.5513. Both backward elimination and stepwise arrived at identical models with eight independent variables (*METHOD6*, forced *CONTRAILS0* inclusion, $R^2 = 0.5698$, $F = 475.26$, $p < 0.0001$). The parameter estimate for *CONTRAILS0* was 6.5173 (standard error = 0.3769, $t = 17.29$, $p < 0.0001$).

The minimal *CONTRAILS0* only model, *METHOD7*, was also statistically significant but exhibited a much lower $R^2 = 0.0973$ ($F = 310.01$, $p < 0.0001$) with a much higher parameter estimate of 9.3605 (standard error = 0.5316, $t = 17.61$, $p < 0.0001$). Among the three, *METHOD6* was selected as preferred. Parameter estimates appear in Table 5.

CONTRAILS: CAUSAL INTERFERENCE

Table 5. Regression and MCMC results

ID	Variable	REGRESSION <i>METHOD6</i> Estimate	<i>p</i> -value	MCMC <i>METHOD9</i> Posterior Mean	95% HPD Interval
0	<i>Intercept</i>	17.5910	<0.0001	17.5903	16.1170 , 19.0984
1	<i>CONTRAILS0</i>	6.5173	<0.0001	6.5195	5.7795 , 7.2552
2	<i>temp</i>	0.5114	<0.0001	0.5112	0.4728 , 0.5493
3	<i>dewp</i>	-0.5473	<0.0001	-0.5471	-0.5857 , -0.5091
4	<i>slp</i>	ELIMINATED	>0.05	NA	NA
5	<i>visib</i>	ELIMINATED	>0.05	NA	NA
6	<i>wdsp</i>	-0.5771	<0.0001	-0.5770	-0.6207 , -0.5318
7	<i>MXSPD</i>	ELIMINATED	>0.05	NA	NA
8	<i>PRCP</i>	ELIMINATED	>0.05	NA	NA
9	<i>p133fog</i>	ELIMINATED	>0.05	NA	NA
10	<i>p134rain</i>	-1.8677	<0.0001	-1.8689	-2.4019 , -1.3160
11	<i>p135snow</i>	-3.9595	<0.0001	-3.9653	-5.7258 , -2.1473
12	<i>p137thun</i>	1.5057	0.0278	1.5053	0.1708 , 2.8489
13	<i>elev</i>	0.0026	<0.0001	0.0026	0.0023 , 0.0029
14	<i>latitudecorr</i>	ELIMINATED	>0.05	NA	NA
15	<i>latitudeabs</i>	ELIMINATED	>0.05	NA	NA
16	<i>temprange</i>	DEP VAR	>0.05	DEP VAR	NA

*Note. Regression parameter estimates then served as MCMC priors

METHOD8 / Naive Regression, Resampling Because *METHOD6* was preferred over the reduced model, only the former was subjected to resampling. For the 1620 runs, the *CONTRAILS0* estimate had mean = 6.6889 (median = 6.6336, range 4.4486, 8.8783). The mean *F* was 150.6304, mean $R^2 = 0.5752$. Each of the 1620 individual runs were statistically significant ($p < 0.0001$). The results of the runs of the various *ntrt* and control to treatment observational ratios are also summarized in Table 6.

Table 6. Resampling results

<i>ntrt</i>	Ratio	Runs	METHOD2 PROP / NO OVERLAP		METHOD4 PROP / OVERLAP		METHOD8 REGRESSION	
			<i>CONTRAILS0</i> ESTIMATE	\DeltaPS	<i>CONTRAILS0</i> ESTIMATE	\DeltaPS	<i>CONTRAILS0</i> ESTIMATE	R ²
136.3	4.8	1620	6.7871	0.0194	6.9654	0.0141	6.6889	0.5752
288	2	180	7.0458	0.0615	7.1010	0.0545	6.5149	0.5750
216	3	180	6.6921	0.0275	6.8047	0.0218	6.5790	0.5743
162	4	180	6.6644	0.0181	6.8320	0.0125	6.6631	0.5750
136	5	180	6.6833	0.0136	6.8779	0.0086	6.7120	0.5746
108	6	180	6.6957	0.0123	6.9054	0.0074	6.7504	0.5741
96	7	180	6.8237	0.0107	7.0324	0.0061	6.7797	0.5747
81	8	180	6.8053	0.0108	7.0089	0.0058	6.7693	0.5753
72	9	180	6.7258	0.0107	6.9667	0.0053	6.7130	0.5768
68	10	180	6.9478	0.0098	7.1593	0.0046	6.7184	0.5771

*Note: Top row is mean for all runs; other rows are means for that resample level

Bayesian

METHOD9 / MCMC, Best Model The best model estimated *CONTRAILS0* as 6.5195 ($\alpha = 0.05$ HPD Interval 5.7795, 7.2552). The model was a normal posterior predictive distribution with normal priors for effects and inverse gamma for variance. Non-informative priors failed to generate a reasonable model, based upon diagnostic plots or Geweke. Therefore, informative priors were set as the estimates from the best model regression, *METHOD6* (Table 5). Variances were set at 100, except elev which was 5×10^{-8} . The MCMC was performed with five blocks: (1) *CONTRAILS0*, (2) *Intercept*, *temp*, *dewp*, *wdsp*, (3) *p134rain*, *p135snow*, *p137thun*, (4) *elev*, and (5) σ^2 . The groups were based on correlations and commonality of data collection.

Acceptance rates ranged from 0.2200 to 0.3040 at the end of the tuning period, 540k burn-in, and 648k sampling. Visually, the diagnostic plots revealed convergence of parameter means, increasingly diminished autocorrelations, and normal posterior density distributions (Table 5, Figure 2). Geweke diagnostic was 0.6480 for *CONTRAILS0*, and ≥ 0.1181 for all others. All were ≥ 0.05 , indicative that the final 50% of runs featured posterior parameter estimates that were not statistically different than of the initial 10%.

CONTRAILS: CAUSAL INTERFERENCE

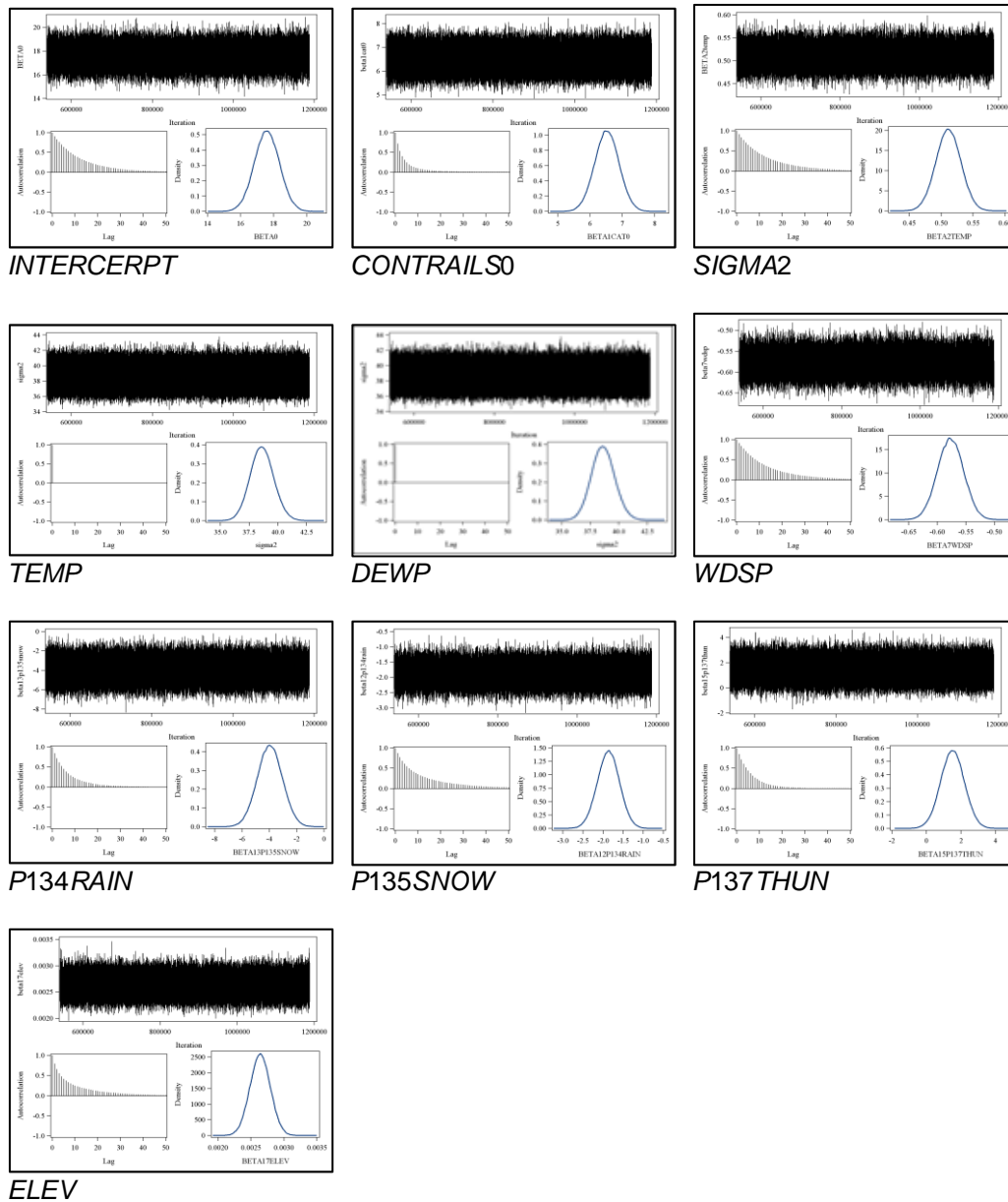


Figure 2. Diagnostic Plots, *METHOD9*

METHOD 10 / MCMC, Minimal Model The minimal model estimated *CONTRAILS0* as 9.3609 ($\alpha = 0.05$ HPD Interval 8.3327, 10.4145). The model was a normal posterior predictive distribution with normal priors for effects and

inverse gamma for variance. The MCMC was performed with three blocks, one for each of β_0 (intercept), β_1 CAT0 (*CONTRAILS0*), and σ^2 . Non-informative priors were used because they proved sufficient.

Acceptance rates were from 0.3528 to 0.3720 for end-tuning period, 72k burn-in, and 360k sampling. Visually, the diagnostic plots also revealed convergence of means, increasingly diminished autocorrelations, and normal posterior density distributions (Figure 3). Geweke diagnostics were all > 0.1700 .

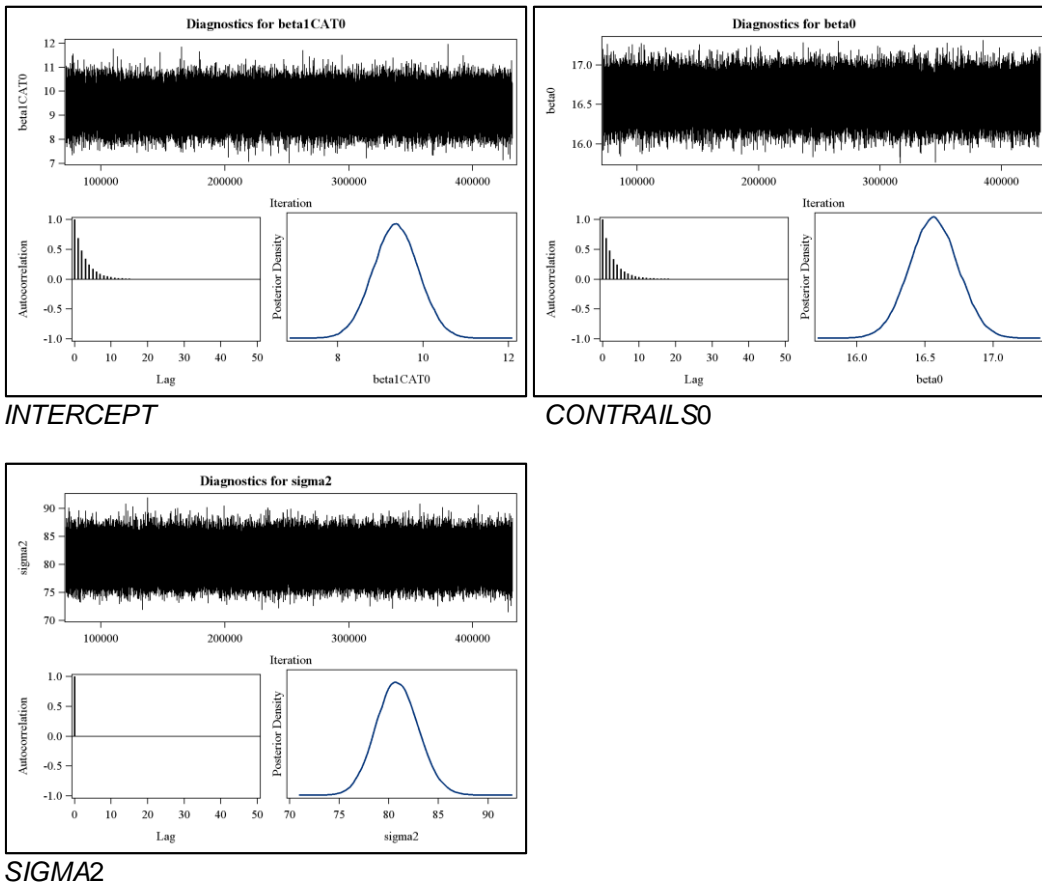


Figure 3. Diagnostic Plots, *METHOD10*

CONTRAILS: CAUSAL INTERFERENCE

CONTRAILS0 Estimate Testing The *CONTRAILS0* estimates from MCMC simulation *METHOD9* analysis revealed that 15.65% were greater than the estimate from *METHOD1*. For any run, the minimum MCMC posterior estimate was 4.8125; the maximum was 8.2761. Thus, all 180k runs were greater than the Travis estimate (Table 7).

Table 7. Comparison of *CONTRAILS0* Estimates

METHOD	TYPE	DESCRIPTION	CONTRAILS0 ESTIMATE	METHOD9 PERCENT >CONTRAILS0
1	PROPENSITY		6.8981	15.65%
2	PROPENSITY	RESAMPLING	6.7871	23.79%
3	PROPENSITY	OVERLAP	6.8931	15.96%
4	PROPENSITY	RESAMPLING & OVERLAP	6.9654	11.79%
5	REGRESSION	FULL MODEL	6.5513	46.77%
6	REGRESSION	BEST MODEL	6.5173	50.41%
7	REGRESSION	MINIMAL MODEL	9.3605	0.00%
8	REGRESSION	BEST MODEL & RESAMPLING	6.6889	32.71%
9	MCMC	BEST MODEL	6.5195	50.15%
10	MCMC	MINIMAL MODEL	9.3609	0.00%
		Travis estimate	3.24	100.00%

Covariate and contrail effect estimate comparisons Of the 14 covariates, the original data had 10 with statistically significant mean differences between the treatment and control groups, as indicated by the bold figures in Table 8. With *METHOD1*, for all covariates, one fails to reject the H_0 that the means in the treatment and control groups are equal.

Table 8. Comparison of Covariates of Original and Propensity Score Matched Data

ID	Variable	ORIGINAL (n _{tt} =2879)				MATCHED (n _{tt} =644)			
		variance	df	t-value	p-value	variance	df	t-value	p-value
1	<i>CONTRAILS0</i>				NA				NA
2	<i>temp</i>	Unequal	505	-9.02	<0.0001	Unequal	595	-0.18	0.8573
3	<i>dewp</i>	Unequal	439	-1.36	0.1746	Unequal	621	0.13	0.8983
4	<i>slp</i>	Unequal	453	-6.26	<0.0001	Equal	642	-0.16	0.8738
5	<i>visib</i>	Unequal	696	8.37	<0.0001	Unequal	518	-0.37	0.7148
6	<i>wdsp</i>	Unequal	750	5.34	<0.0001	Unequal	517	0.01	0.9933
7	<i>MXSPD</i>	Unequal	2614	6.60	<0.0001	Unequal	326	0.74	0.4586
8	<i>PRCP</i>	Unequal	455	1.38	0.1697	Unequal	582	-0.40	0.6861
9	<i>p133fog</i>	Unequal	366	-3.89	<0.0001	Equal	642	0.31	0.7530
10	<i>p134rain</i>	Equal	2877	2.41	0.0159	Equal	642	-0.77	0.4431
11	<i>p135snow</i>	Unequal	970	4.13	<0.0001	Unequal	578	0.58	0.5635
12	<i>p137thun</i>	Unequal	351	-3.22	0.0014	Equal	642	-0.79	0.4307
13	<i>elev</i>	Unequal	448	-0.39	0.6951	Unequal	634	-1.60	0.1106
14	<i>latitudecorr</i>	Equal	2877	-0.02	0.9859	Equal	642	-1.33	0.1845
15	<i>latitudeabs</i>	Unequal	518	6.50	<0.0001	Unequal	584	0.37	0.7111
16	<i>TEMPRANGE</i>				NA				NA

*Note. Differences that are statistically significant at $\alpha = 0.05$ are in bold

The *CONTRAILS0* estimate from PS matching using all observations without overlap was 6.8981. This was a statistically significant difference from the Travis estimate ($t = 5.26, p < 0.0001$).

Except for the minimal models (*METHOD7*, *METHOD10*), the contrails effect estimate within the 95% confidence interval of *METHOD1*, and therefore did not represent statistical difference. Due to its simplicity, *METHOD1* was preferred over the other causal inference methods; due to covariate equalities, it was preferred over the regression and MCMC methods.

Omnibus distribution tests

Distributional differences were tested by three omnibus tests, Kolmogoroff-Smirnoff (KS), Cramér-von Mises (CM), and oando (Barron, 2007). The control and treatment group distributional differences were statistically significant for KS for the analysis dataset ($n_{trt} = 322, n_{ctl} = 2557, D = 0.4412, p < 0.001$) and propensity matched data subset ($n_{trt} = n_{ctl} = 322, D = 0.3571,$

CONTRAILS: CAUSAL INTERFERENCE

$p < 0.001$). CM also indicated statistical significance for both datasets (CM = 25.2482, $p < 0.001$ and CM = 8.6263, $p < 0.001$).

Oando performed with resampling yielded inconclusive results. For 4000 resampling runs, the mean p -value = 0.1988, 32.73% p -value ≤ 0.05 , and 11.20% p -value > 0.50 . For 180 resampling runs, the matched dataset had mean p -value = 0.2871, 22.22% p -value ≤ 0.05 , and 20.56% p -value > 0.50 .

Due to the definition of oando which weights by the rank of the gap from the prior observation, the result is possibly reflective of a non-homogenous range. In response, the analysis dataset was partitioned according to *temprange* rank. The low and high ends were curves; the middle was linear. *REGION1* was defined as the union of the low (*REGION1A*) and high (*REGION1B*); *REGION2* was defined as the middle (Figure 4).

The *temprange* difference between control and treatment represented the *CONTRAILS0* estimate. For the entire analysis dataset, the union of *REGION1* and *REGION2*, the *temprange* difference was 9.3605. For mid-*tempranges* of 12.20 to 20.22, the *CONTRAILS0* estimate for *REGION2* was not statistically different from zero. However, for *REGION1* the contrails effect estimate was 11.4521 (p -value < 0.0001).

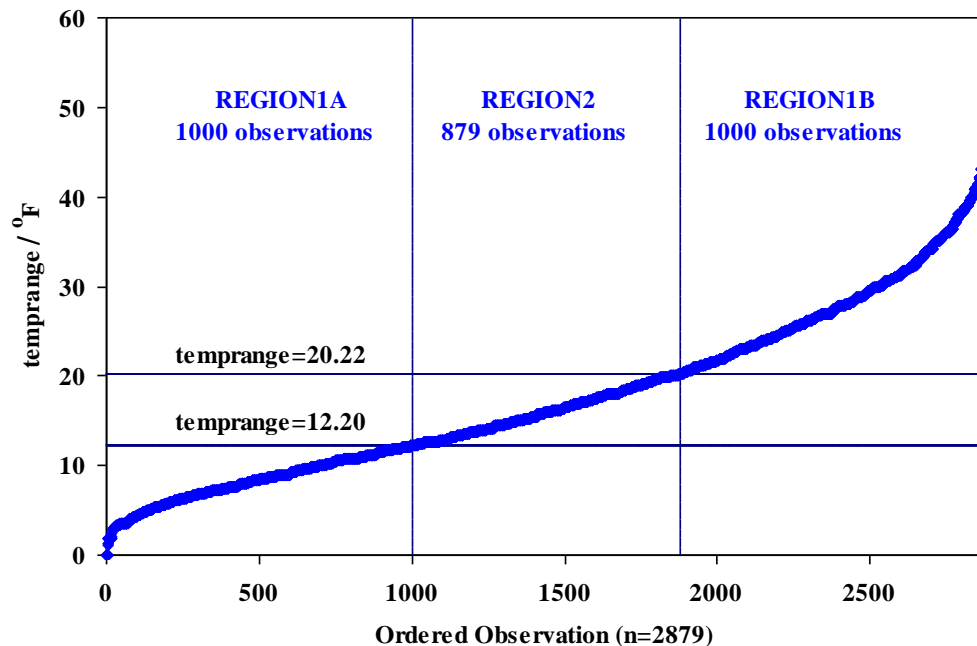


Figure 4. Partition of Ordered Observations of Analysis Dataset

Conclusion

The contrail effect was estimated to equal a 6.8981°F decrease in the daily temperature range at ground level on planet Earth using propensity matching, *METHOD1*. This result was statistically different from the Travis estimate of 3.24°F.

Although rarely studied, daily temperature range does impact animal populations and population dynamics (Viterbi, et al., 2012). Smaller daily temperature ranges have been shown to decrease the black grouse bird population in Italy (Viterbi, et al., 2012) as well as to influence Moluccan Woodcock population density in Indonesia (Eden, et al., 2013). The impact upon other species may also be significant (Eden, et al., 2013).

In pursuit of fuel economy, modern engines sport a greater efficiency of propulsion. However, aircraft equipped with such engines generate contrails starting at lower altitudes (Schumann, et al., 2000), and up to higher altitudes (Schumann, 2000). More persistent contrails could shadow even more of the Earth's surface than the 16% EPA estimation (EPA, 2000).

Analysis variables were solely based upon the NCDC datasets. Other data might have been useful, for example, temperature and other atmospheric measurements taken at altitudes at 25,000 feet; NCDC measures only at ground level. The restriction to data from a single source obviated the need to judge relative reliability of different databases, measurement devices, and data collection procedures.

Two omnibus tests, Kolmogoroff-Smirnoff and Cramér-von Mises, confirmed distributional differences between treatment and control groups, supportive of the propensity score matching results. The third, oando, revealed that the data might be an amalgam of two regions, center and extremes. Future explorations could introduce an indicator variable reflecting such a partition, or, fractionate into individual analyses.

The correlation between daily mean temperature and *CONTRAILS0* of 0.1267 was consistent with an association of higher mean temperatures at ground level with absence of contrails. This was in agreement with the minority; most prior studies have indicated a net warming effect, but inconclusively (Mims, Chambers & Oostra, n.d.). The NCDC data calculates its reported daily mean temperatures based upon the actual operating hours for that specific station (Lott, 2010). Mean temperature theoretically might also be defined as the mean of 24 hourly readings, or many other possible variants. Alternatively, the median measurement might be a reasonable reflection of central tendency. These

CONTRAILS: CAUSAL INTERFERENCE

considerations could cloud conclusions. Because the contrails effect upon daily mean temperature was not the focus of this analysis, techniques employed were not aimed at obtaining such an estimate. Therefore, although interesting, any inferences regarding daily mean temperature are merely ancillary.

Author Contributions

The seminal concept of subjecting contrail-related data to causal inference was by J.H.B., and subsequently developed and discussed by J.H.B. and D.S.B. The logistic regression was coded in SAS by J.H.B. The remainder of the research, SAS coding, and all writing was performed by D.S.B.

References

- Barron, D. S. (2007). *Kolmogoroff-Smirnoff Enhancement* (Master's thesis). Retrieved from ProQuest Dissertations and Thesis Full Text. (1451232).
- Barron, D. S. (Photographer). (2013). *Contrail 33* [photograph], Personal collection.
- Bauer, P., Read, A., & Johnson, P. (2000). *The gravity geophysical method*. Retrieved from http://geoinfo.nmt.edu/geoscience/projects/astronauts/gravity_method.html
- Bristow, K. L & Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, 31(2), 159-166.
doi:10.1016/0168-1923(84)90017-0
- Cottee-Jones, H. E. W., Mittermeier, J. C., & Redding, D. W. (2013). The Moluccan Woodcock *Scolopax rochussenii* on Obi Island, North Moluccas, Indonesia: a 'lost' species is less endangered than expected. *Forktail* (29), 88-93.
- Gelman, A. & Hill, J. (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.
- Gray, D. M. & Male, D. H. (1981). *Handbook of snow: Principles, processes, management and use*. Toronto, Canada: Pergamon Press.
- Lott, N. (2010). The National Climatic Data Center. *Federal climate complex: Global surface summary of day data, version 7* [Data file]. Retrieved from <ftp://ftp.ncdc.noaa.gov/pub/data/g sod/readme.txt>

Meerkotter, R., Schumann, U., Doelling, D. R., Minnis, P., Nakajima, T., & Tsushima, Y. (1999). Radiative forcing by contrails. *Annales Geophysicae*, 17(8), 1080-1094. doi:10.1007/s00585-999-1080-7

Mims, F. M., Chambers, L. & Oostra, D. H. (n.d.). *Contrail studies*. Retrieved from http://mynasadata.larc.nasa.gov/science_projects/contrail-studies/

Moore, M. (Director & Writer). (2004a). *Fahrenheit 9/11* [Motion picture]. United States: Fellowship Adventure Group (presented by), Dog Eat Dog Films, Miramax Films (uncredited).

Moore, M. (2004b). *What Fahrenheit 9/11 says about the Saudi flights out of the country after September 11*. Retrieved from <http://www.michaelmoore.com/words/fahrenheit-911-facts/what-fahrenheit-911-says-about-the-saudi-flights-out-of-the-country-after-september-11>

The National Climatic Data Center. (2010). *Integrated Surface Data, version 7* [Data files]. Available from <ftp://ftp.ncdc.noaa.gov/pub/data/g sod/>
National Retail Federation (2013). *FAQs 4-5-4 calendar*. Retrieved from http://www.nrf.com/modules.php?name=Pages&sp_id=392

Schumann, U. (2000). Influence of propulsion efficiency on contrail formation. *Aerospace Science and Technology* 4(6), 391–401. doi:10.1016/S1270-9638(00)01062-2

Schumann, U., Busen, R., & Plohr, M. (2000). Experimental test of the influence of propulsion efficiency on contrail formation. *Journal of Aircraft*, 37(6), 1083-1087. doi:10.2514/2.2715

Schumann, U. (2005). Formation, properties and climatic effects of contrails. *Comptes Rendus Physique*, 6(4-5), 549-565. doi:10.1016/j.crhy.2005.05.002

Travis, D. J., Carleton, A. M., & Lauritsen, R. G. (2002). Climatology: Contrails reduce daily temperature range, brief communications. *Nature*, 418, 601. doi:10.1038/418601a

United States Environmental Protection Agency. (2000). *Aircraft Contrails Factsheet*. EPA430-F-00-005, September 2000 Air and Radiation (6205J).

Viterbi, R., Imperio, S., Alpe, D., Peverelli, V.B., & Provenzale, A. (2012). Climatic control and population dynamic of black grouse in the Western Italian Alps. *Istituto di Scienze dell'Atmosfera e del Clima, CNR, Relazioni finale*. Retrieved from http://www.regione.piemonte.it/agri/area_tecnico_scientifica/osserv_faun/progetti/dwd/alcotra/relFinali/1T.pdf

CONTRAILS: CAUSAL INTERFERENCE

Wendler, G. & Stuefer, M. (2002). *Improved contrail forecasting techniques for the subarctic setting of Fairbanks, Alaska*. Geophysical Institute University of Alaska, University of Alaska Fairbanks Special Report UAG R-329.