Skillful Seasonal Forecasts of Winter Disruption to the U.K. Transport System

ERIKA J. PALIN, ADAM A. SCAIFE, EMILY WALLACE, EDWARD C. D. POPE, ALBERTO ARRIBAS, AND ANCA BROOKSHAW

Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received 1 April 2015, in final form 24 August 2015)

ABSTRACT

The impacts of winter weather on transport networks have been highlighted by various high-profile disruptions to road, rail, and air transport in the United Kingdom during recent winters. Recent advances in the predictability of the winter North Atlantic Oscillation (NAO) at seasonal time scales, using the Met Office Global Seasonal forecasting system, version 5 (GloSea5), present a timely opportunity for assessing the long-range predictability of a variety of winter-weather impacts on transport. This study examines the relationships between the observed and forecast NAO and a variety of U.K. winter impacts on transport in the road, rail, and aviation sectors. The results of this preliminary study show statistically significant relationships between both observed and forecast NAO index and quantities such as road-accident numbers in certain weather conditions, weather-related delays to flights leaving London Heathrow Airport, and weather-related incidents on the railway network. This supports the feasibility of the onward goal of this work, which is to investigate prototype seasonal forecasts of the relative risk of occurrence of particular impacts in a given winter for the United Kingdom, at lead times of 1–3 months. In addition, subject to the availability of relevant impacts data, there is scope for further work to make similar assessments for other parts of Europe and North America where the NAO has a strong effect on winter climate.

1. Introduction

The U.K. transport system exhibits a variety of weather sensitivities. Examples include the effects of very hot weather, such as buckling of rails and degradation of road surfaces; flooding (of fluvial, surface water, or groundwater origin) of assets and routes; lightning impacts on electronic systems such as those used in railway signaling; failure of locomotives as a result of snow ingress to traction equipment; wind throw of trees and other objects onto roads and railways; the influence of prevailing weather conditions on the occurrence of road traffic accidents; increased need for deicing of aircraft and gritting of highways in extremely cold weather; safety-related closure of airport runways during snowy, icy, or foggy conditions; and impacts on marine transport operations from high waves during stormy conditions. In recent years the U.K. transport system has experienced a number of high-profile disruptions from winter weather; Table 1 summarizes these weather events and a selection of their impacts on various transport modes.

The transport sector is accustomed to using weather-forecast products to inform operational planning on short time scales (e.g., RSSB 2013). In recent years the sector has also begun to develop its knowledge at the climate-change time scale, typically using new insights thus gained to inform climate-change adaptation plans or the direction of future sector-relevant climate-change impacts studies (e.g., Palin et al. 2013). Although the potential value of information at seasonal time scales has previously been highlighted (Juga et al. 2012), until recently there has not been sufficient skill in operational seasonal forecasts to make it feasible to use them in the sector, despite there being a strong appetite for this, as evidenced by a U.K. government report on the impacts of the December 2010 weather on transport:

Better medium- and long-range weather forecasting would assist transport providers and others in planning to deal with the effects of severe winter weather. For example, it would give transport operators the opportunity...
to warn passengers of when contingency timetables would be likely to be needed and to get snow and ice clearing equipment into position. £10 million would be a small price to pay for improving the Met Office’s long-range forecasting capability, given the cost to the UK economy of transport disruption due to severe winter weather (House of Commons Transport Committee 2011, p. 9).

In this paper we discuss new developments in the Met Office Global Seasonal forecasting system (GloSea;...
that could permit the development of risk-based seasonal forecasts of U.K. winter transport disruption. We use examples of impacts on rail, road, and aviation to illustrate the relationships on which these forecasts could be built.

2. Predictability of winter conditions at seasonal time scales

The North Atlantic Oscillation (NAO) is a major driver of northern European winter climate (Hurrell 1995). It is typically quantified with an index, defined as the difference between the mean sea level pressure in Iceland and that in the Azores. Positive NAO index values are typically linked with wetter, warmer, stormier conditions in northern Europe, whereas negative values are typically linked with calmer and colder conditions in northern Europe. By way of example, the NAO index during the extremely cold winter of 2009/10 was the most negative on record (Fereday et al. 2012).

Met Office seasonal forecasts are currently produced operationally using version 5 of GloSea (GloSea5; MacLachlan et al. 2014). The forecast system uses the Hadley Centre Global Environmental Model, version 3 (HadGEM3), with atmospheric resolution of 0.83° longitude × 0.55° latitude, 85 quasi-horizontal atmospheric levels, and an upper boundary at 85 km near the mesopause (Walters et al. 2014). The ocean component has a resolution of 0.25° globally in both latitude and longitude, with 75 quasi-horizontal levels. The initial conditions used to initialize the retrospective forecasts discussed here are taken from ERA-Interim observational reanalyses (Dee et al. 2011) (atmospheric and land surface data) and the FOAM data-assimilation system (Blockley et al. 2014) (global ocean and sea ice concentration data).

With the advent of GloSea5, there has been a step change in the ability of the seasonal-forecast system to predict the NAO index during winter [December, January, and February (DJF)] (Scaife et al. 2014), with similar progress being demonstrated for the related Arctic Oscillation (the hemispheric equivalent of the NAO) in other seasonal-forecast systems (Riddle et al. 2013; Kang et al. 2014). As such, there is now scope for using NAO-index predictions as a basis for risk-based seasonal forecasts, on the condition that appropriate relationships can be found between the forecast NAO index and metrics for transport-system impacts.

Given the predictability of the NAO index and the link between the NAO and frequency of extreme winter events (Scaife et al. 2008; Kenyon and Hegerl 2008), here we use the NAO as a proxy predictor for weather events with impacts on transport. Because most of the European winter predictability in GloSea5 originates from the prediction skill of the NAO (Scaife et al. 2014), we use this large-scale circulation measure as a source of forecast information. Indeed, to the extent that observed extreme events are governed by large-scale circulation, it may be the case that even in a perfect forecast system the approach of using the NAO as a proxy predictor could yield the most skillful forecast, because counting the number of weather events inevitably introduces some statistical noise (see later discussion).

a. Meteorological data used in this analysis

We used both observed and modeled NAO indices in this analysis. Observed NAO data were computed from HadSLP2r, a combination of land and marine pressure observations from 1850 to 2004. HadSLP2r is a near-real-time, updated form of the original (HadSLP) dataset, which uses NCEP–NCAR reanalysis fields, adjusted to account for the differences between HadSLP and the NCEP–NCAR products (Allan and Ansell 2006).

Modeled NAO values were computed from GloSea5 hindcast runs (MacLachlan et al. 2014), an existing set of retrospective forecasts, for the 20 winters from 1992/93 to 2011/12, with 24 forecast members per winter. Forecasts for each winter start from dates centered on the beginning of November, and forecasts from the same date differ only by stochastic physics. We will refer to NAO-index values from the GloSea5 hindcast runs as “forecast NAO index” values in this paper.

For ease of comparison, we have normalized each dataset (observed and modeled NAO) by its standard deviation to provide a standardized index. While the interannual correlation is high, the forecast ensemble-average NAO values have small amplitude (Scaife et al. 2014; Eade et al. 2014). This suggests that our method of using the NAO as a predictor of impacts is currently preferable to using actual forecast extreme-weather events, which will show only a small change in frequency. Figure 1 shows a plot of observed NAO index against that forecast by GloSea5. The correlation coefficient is 0.62, but this is dependent on the forecast ensemble size, indicating potentially greater skill for a larger ensemble size (Scaife et al. 2014). In particular, the forecast system reproduces very well the relatively extreme NAO indices of the high-impact winters of 2009/10 and 2010/11 (both negative; cold-related impacts) and 2011/12 (positive; rain/wind-related impacts) that are discussed above. The winter of 1999/2000, also with a strongly positive NAO index, did not see such widespread impacts in the United Kingdom—although several high-impact storms affected other parts of Europe, including “Lothar” and “Martin,” both of which had notable effects on transport in France and Switzerland.
In addition to the correlations presented in Fig. 1, the forecast system also frequently captures the large-scale pattern in the DJF NAO [plots showing hindcast ensemble mean SLP anomaly vs observed SLP anomaly can be seen in Scaife et al. (2014), their Fig. S1].

In this study our primary focus is on the impacts summarized over a whole winter. It is clear from the above that the spatial distribution of impacts during a given winter may depend on the paths of particular weather systems within a winter, with different winters yielding different levels of impacts in the United Kingdom. This uncertainty in the spatial variability of impacts is inherent in the multiyear assessment we have performed. We do briefly touch on the cases of individual winters in our statistical analyses (see section 2c and subsequent specific cases).

b. Impacts data used in this analysis

The impacts data we have used are discussed in sections 3, 4, and 5 for aviation, road, and rail impacts, respectively. In each case we summarize the available impact data as a dataset with one value per winter and perform linear-regression analysis (including significance testing and discussion of unusual data points) to assess relationships between the NAO and transport impact. There are limitations associated with using a linear-regression approach (such as the fact that, in reality, for very extreme NAO indices the impacts could saturate), but we have focused on the linear relationship in this study because of the sometimes limited nature of available impacts data (e.g., short time series) and the preliminary state of seasonal forecasting in this area.

In many cases there is an obvious a priori sign of correlation anticipated between NAO index and impact—for example, deicing metrics would be expected to correlate negatively with the NAO index (more deicing expected in lower NAO-index conditions). In these cases we have performed one-tailed significance tests. Other cases are less clear and there is not an a priori expectation—for example, a metric of disruption may simply be described as “weather related” and may contain contributions from circumstances ultimately related to either negative or positive NAO conditions. In these cases we have used a two-tailed significance test. We have also used standard methods to assess the (lag 1) autocorrelation in all of the impact metric time series discussed in this paper, the autocorrelation for the NAO time series having already been assessed (Scaife et al. 2014). The approaches we have taken to characterize the impacts are discussed as they arise below.

c. Statistical analysis of correlations

Aside from demonstrating a correlation between meteorological conditions and impacts on the U.K. transport network, we also highlight several years in which either or both of the NAO index and impact are unusual in the context of the available data. These unusual years can influence the strength of the correlation along with the linear-regression model’s slope and intercept, and therefore it is important to investigate their
statistical properties in relation to the whole sample. To achieve this, it is standard to identify outliers and points with high leverage and influence (e.g., Montgomery and Peck 1992). Outliers are defined as points that have residuals (i.e., displacement from the predicted value in the y direction) greater than a critical value defined by the properties of the sample. For this purpose, we employ the Bonferroni test to determine whether the largest residual in each dataset is an outlier at the 95% confidence level (Montgomery and Peck 1992).

Leverage is quantified by the amount that an observation’s value on the predictor variable (here, the NAO index) differs from the mean of the predictor variable. In other words, high leverage points are those that lie away from the main cluster of data points. They are potentially important since they tend to have small residuals and can have a significant effect on the regression line fit. In general, a given point’s leverage is considered to be high if it exceeds a putative threshold of $\frac{2}{n} \times k$ (e.g., Montgomery and Peck 1992), where $n$ is the sample size and $k$ is the number of fitted parameters in the regression model (i.e., $k = 2$ in this case).

Influential points are defined as having the largest impact on the magnitude of the correlation. This influence can be quantified using the Cook’s distance, which measures the difference in the regression estimates when each observation is omitted in turn. The Cook’s distance of an individual point is generally considered to be large if it exceeds a threshold of $\frac{2}{n}$ (e.g., Montgomery and Peck 1992), but there is no universally agreed threshold, and other values may be equally applicable. For this reason, the outcomes of these tests must be treated with some caution.

The results from these statistical assessments are discussed for each transport subsector below. Note that in this paper we will use the term “unusual” to refer specifically to years that are statistically unusual according to the assessments outlined above.

3. Impacts on aviation

To examine aviation disruption, we used data supplied by British Airways (BA) for delays at London Heathrow Airport (LHR). Data were supplied for the number and percentage of BA flights departing LHR that were delayed by weather and for the length of weather-related delays. A further dataset was supplied for the number of BA aircraft deiced at LHR during winter. In both cases, BA uses as its definition of winter the International Air Transport Association “northern winter” season (IATA 2014), which corresponds to the period over which European Union countries do not use daylight saving time (late October–early March); this is, of course, an extension of the conventional meteorological definition. Note that, despite the longer coverage period for the impact data, we are still using the DJF NAO as the proxy predictor.

The deicing data showed evidence of an increasing trend. Information from BA suggested that their deicing policy had not seen any major changes during the period of data availability, and as such it was assumed that the trend arose from a general increase in LHR’s capacity. The number of BA flights departing LHR also shows an upward trend that supports this assumption; the deicing data were therefore linearly detrended before use in this analysis. Figure 2 shows the relationship between the NAO index and the three different aviation impacts outlined above.

There are clear relationships between each impact and the NAO index. In each case, the slope of the relationship is negative, with larger impacts at lower NAO values. For two of the three impact metrics the nature of this slope could not be predicted a priori (i.e., for number and percentage of delays it is not clear whether strongly positive NAO conditions or strongly negative NAO conditions should have the most impact). It turns out that it is the negative NAO conditions that prevail in these relationships, in accordance with that for the number of aircraft deiced, where the negative slope could be predicted a priori (see discussion in section 2b above). Using a one-tailed test, we found the correlation between aircraft deiced at LHR and observed NAO index to be significant at the 95% level; the relationship for the forecast NAO index was not significant. Using a two-tailed test, the correlations of length and percentage of weather-related delays at LHR were found to be significant for both the observed and forecast NAO index. The outcomes of the significance testing are summarized later (in Table 6, below).

The results of the statistical analysis on data points with unusual NAO indices or aviation impacts are shown in Table 2. In all cases, 2009/10 has the highest leverage since it lies farthest from the main cluster of points. Furthermore, the leverage of 2009/10 is considered to be large for each aviation-impact metric since it exceeds the threshold defined above. Given this, we might expect 2009/10 to be the most influential point on the correlation, as quantified by the Cook’s distance, but this is not the case. For both the percentage and length of weather-related delays at LHR, the largest residuals and most influential points both occur in 2006/07, which exhibits impacts that are comparable to 2009/10 despite a moderately positive NAO. However, 2006/07 is not classified as an outlier at the 95% confidence level. Taking this information together, it appears that, despite its large
leverage, the presence of 2009/10 in the dataset does not influence the correlation disproportionately. Instead, these results indicate that 2009/10 was broadly consistent with the expected impact of an extremely negative NAO index.

For the number of BA aircraft deiced, the largest residual occurs for 2012/13, which is also the most influential point. It seems plausible that this is a consequence of the extended period of intensely cold conditions in the spring of 2013, but, again, further investigation would be required to verify this. Note also that, across these three aviation-impact metrics, the year with the largest residual is also the most influential point. At present, it is not clear whether this phenomenon has arisen by chance as a result of limited data availability or whether it is suggestive of particular sensitivities for these aviation-impact metrics. In contrast, results for the road and rail sectors (discussed below) suggest that the largest residual is not always the most influential point. For both of these reasons, in future work it will be important to understand the origin of large residuals (e.g., the 2006/07 anomaly) so as to better understand the sources of noise in the relationship between weather and impacts. We also found a moderate lag-1 autocorrelation ($0.4–0.5$) in the “percentage of weather-related delays” and “length of weather-related delays” time series, suggesting that

![Figure 2](image_url)

**Fig. 2.** Relationship between (left) observed or (right) forecast NAO index and a selection of winter aviation impacts at LHR: (top) percentage of November–March delays due to weather, (middle) length of November–March delays due to weather, and (bottom) number of BA aircraft (wide bodied plus narrow bodied) deiced at LHR. The data for number of aircraft deiced have been linearly detrended. For each panel, the straight line indicates the linear-regression fit and the dashed curves indicate the 95% confidence intervals on the fit. Labels indicate the December year of the winter. Correlation coefficients and $p$ values on the fits are specified at the top right of each panel.

<table>
<thead>
<tr>
<th>Impact metric</th>
<th>Residual</th>
<th>Leverage</th>
<th>Cook’s distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of weather-related delays to flights departing LHR</td>
<td>2006/07</td>
<td>2009/10</td>
<td>2006/07</td>
</tr>
<tr>
<td>Length of weather-related delays at LHR</td>
<td>2006/07</td>
<td>2009/10</td>
<td>2006/07</td>
</tr>
<tr>
<td>Aircraft deiced at LHR</td>
<td>2012/13</td>
<td>2009/10</td>
<td>2012/13</td>
</tr>
</tbody>
</table>

**Table 2.** Summary of winters with the largest residual, leverage, and Cook’s distance for the aviation-impact data. Years that are statistically significant, according to the criteria presented in section 2c, are shown in boldface font.
further caution should be applied in interpreting correlations involving these data.

4. Impacts on the road sector

We used two datasets to assess impacts on the road sector: the “STATS19 Road Accident” dataset (Department for Transport 2014) and the amount of salt used by the Highways Agency (now known as Highways England) for deicing purposes. The STATS19 dataset includes a wealth of information about road accidents involving human injury or fatality. In the STATS19 dataset the exact cause of a specific accident is not attributed—rather, a range of relevant parameters about the circumstances of the accident are recorded. In this analysis we examine only weather-related factors and we do not attempt to consider other factors such as human error, vehicle fault, and so on. The weather conditions at the time of an accident may be recorded as 1) fine with no high winds, 2) raining with no high winds, 3) snowing with no high winds, 4) fine plus high winds, 5) raining plus high winds, 6) snowing plus high winds, 7) fog or mist, 8) other, 9) unknown, and 10) data missing or out of range.

Of these classifications, we assessed the correlation between the NAO index and the classifications most clearly associated with either positive or negative NAO conditions, which are raining plus high winds (positive) and snowing with no high winds (negative), shown in boldface font in the list above. Also calculated were combinations of these classifications, giving data for “all accidents in windy conditions” (fine plus high winds, raining plus high winds, and snowing plus high winds); “all accidents in snowy conditions” (snowing with no high winds and snowing plus high winds); “all accidents in wet conditions” (raining with no high winds and raining plus high winds); and “all accidents in rainy conditions” (raining with no high winds and raining plus high winds). All data were expressed as a percentage of the total winter accidents, because there appears to be a downward trend in the total number of road accidents beginning in about 1999 (not shown). This trend could be due to factors that are not weather related, such as improvements in road-safety legislation or improvements in safety requirements for car manufacture.

Data were provided by the Highways Agency for the total annual salt usage (2003/04 onward) in metric tons on the road network for which it is responsible (motorways and major trunk roads in England). Data at a finer time resolution were not available, and so correlations have been performed with the annual data, which we can assume are mainly due to winter-season usage.

Figure 3 shows the relationships between observed and forecast NAO indices and the various road-impact metrics. For the accident categories of raining plus high winds and snowing with no high winds, the sign of the fitted relationship is opposite for the two different categories, and the signs of the correlations are as expected a priori, given the impact of the NAO on winter extremes (Scaife et al. 2008). The quality of the fit is comparable for both of these metrics, with the relationships found to be significant at the 95% level (one-tailed test) for the correlations with both the observed and forecast NAO, except for the correlation between snowing plus high winds and forecast NAO.

For the combined accident categories (see Fig. A1, below) the correlations are also still fairly strong. Again, although these combined categories include contributions that may have an element of association with both positive and negative NAO conditions, the signs of the correlations that we might expect from a priori considerations are preserved (i.e., negative correlations of NAO index with accidents in snowy conditions, but positive correlations of NAO index with accidents in wet conditions and in windy conditions). For these categories we used a two-tailed test and found that only the correlations between the observed NAO and all accidents in snowy conditions and between the observed NAO and all accidents in windy conditions were significant according to this test.

For road-salt usage, the relationship is again as expected, with higher impacts at lower NAO values. Using a one-tailed test, we found that the correlation was significant only with the observed NAO. (The outcomes of significance testing are given in Table 6, described in more detail below.)

The statistical analysis of unusual points in the road-impact data is shown in Table 3. Again, in all cases 2009/10 has the highest leverage, with values exceeding the critical threshold. For each of the impact metrics, the years with the largest residuals are 1995/96, 1989/90, and 2006/07, respectively. However, only the 1989/90 residual for accidents in wet and windy weather is large enough for it to be considered an outlier at the 95% confidence level. The most influential points, as measured by the Cook’s distance are 2009/10, 1989/90, and 2006/07, respectively. From this sample, 1989/90 appears to be an anomalous winter, with the magnitude of the impacts far exceeding what would be expected given the relatively small positive value of the average NAO index. One possible explanation is that many of the accidents could be associated with the single “Burns’ Day storm,” which affected much of the United Kingdom on 25 January 1990 (Rowe 1990), causing widespread travel disruption. As with the aviation data, findings such as this suggest that a significant fraction of the scatter in the correlations could be a result of variability in the NAO index throughout the winter, along with the impact of notable events that are not strongly correlated with the NAO. Again, further investigation in the future would be required to assess the extent of this effect. Also, there were no significant lag-1 autocorrelations in any of the road-impact-metric time series.
5. Impacts on the rail sector

Information about delays and incidents on the railway network in Great Britain is collated by Network Rail to satisfy the “Schedule 8” performance regime (e.g., Network Rail 2013). Delays may be attributed to a wide variety of causes, such as track faults, signaling failures, fires, vandalism, and bridge strikes. As well as these causes, there are several classifications that are deemed by Network Rail to be weather related. Incidents are recorded into the “TRUST” database, with each entry including information such as date, cause, description, affected route, number of “delay minutes,” and delay cost.

For this analysis we used a dataset supplied by Network Rail’s National Performance Analysis team. The dataset is effectively a summary over the TRUST dataset, including only the incident causes that were deemed by Network Rail to be weather related. Data are available from 2004 onward (a comparatively short time series), with monthly totals for incident counts, train-affected counts, and delay minutes caused for each of England, Wales, and Scotland. (The full list of incident causes deemed by Network Rail to be weather related can be found in Table A1 of appendix A.)

On examining the data, the incident categories in Table 4 were considered to be relevant to this analysis. We excluded some of the categories that were listed in appendix A, as follows:

- categories related to the leaf-fall season are excluded because the weather sensitivities associated with leaf

Table 3. As in Table 2, but for the road-impact data.

<table>
<thead>
<tr>
<th>Impact metric</th>
<th>Residual</th>
<th>Leverage</th>
<th>Cook’s distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of serious road accidents in Great Britain in snowy and calm weather</td>
<td>1995/06</td>
<td>2009/10</td>
<td>2009/10</td>
</tr>
<tr>
<td>Percent of serious road accidents in Great Britain in wet and windy weather</td>
<td>1989/90</td>
<td>2009/10</td>
<td>1989/90</td>
</tr>
<tr>
<td>HA annual salt usage</td>
<td>2006/07</td>
<td>2009/10</td>
<td>2006/07</td>
</tr>
</tbody>
</table>
fall are very complex and are difficult to disaggregate from one another,
• lightning-related causes were omitted because of their rarity in winter, and
• it was assumed that, during winter, categories referring to “extreme heat or high wind” could be classed as due to high wind only.

The majority of incident classes in Table 4 are infrastructure related, but the “701D MW” classification gives data for impacts on train and rolling-stock equipment (hereinafter T&RS). As such, this category was separated out from the others. The “type” column of Table 4 indicates whether the incidents are likely to be related to negative or positive NAO conditions or whether this such a relation cannot be determined from the incident reason. DJF incident counts were computed over England, Wales, and Scotland as a whole.

Figure 4 shows the weather-related incident counts as a function of observed and forecast NAO indices. Relationships are presented for infrastructure incidents (all categories except 701D MW), T&RS incidents (701D MW only), and as a total across all categories in Table 4. All of these correlations are negative, and the impactful winters of 2009/10 and 2010/11 are clearly visible as prominent extremes, as expected. The correlations of infrastructure-incidents and all-incident categories with observed NAO were found to be significant at the 95% level; all other correlations were found not to be significant (using a two-tailed test). (Table 6, described later, summarizes the significance testing outcomes.)

We also separated the data in Table 4 into positive and negative NAO categories (not shown) but found the correlation for positive NAO-only impacts to be poor; this result is initially surprising given the high-profile effects of
flooding seen on the railway network in recent years. During previous analysis of the railway-impact data for other purposes, however, we have noted that other weather-related impacts are recorded under delay codes that are not included in the Network Rail dataset of weather-related-incident causes. For example, a delay category for “earthslip/subsidence/breached sea defences” exists that could easily include some delays and incidents that were ultimately caused by weather, but this code is not included in Network Rail’s list in appendix A. It would be complex to analyze some of these other delay categories, though, because they do not solely reflect weather-related incidents. It is likely that some impacts of recent wet and/or stormy winters are not well represented in this dataset, which covers only delay minutes and “deemed delay minutes” due to partial and full cancellations of train services. Particularly severe weather can result in prolonged line closures, for example, which would be recorded in a different dataset. A wider analysis, incorporating other datasets for weather-related impacts on the railway network, would therefore be valuable.

Table 5 shows the statistical analysis of unusual points in the rail-impact data. In all cases the winter with the highest leverage is 2009/10. As was the case for both aviation and road impacts, the leverage values for 2009/10 exceed the critical threshold. The largest residuals for each of the impact metrics occur in 2010/11, but only the residual for incidents affecting T&RS is large enough for it to be considered an outlier at the 95% confidence level.

**Table 5.** As in Table 2, but for the rail-impact data.

<table>
<thead>
<tr>
<th>Impact metric</th>
<th>Residual</th>
<th>Leverage</th>
<th>Cook’s distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delays on the railway network in Great Britain affecting infrastructure</td>
<td>2010/11</td>
<td>2009/10</td>
<td>2010/11</td>
</tr>
<tr>
<td>Delays on the railway network in Great Britain affecting T&amp;RS</td>
<td>2010/11</td>
<td>2009/10</td>
<td>2009/10</td>
</tr>
<tr>
<td>Delays on the railway network in Great Britain: all categories</td>
<td>2010/11</td>
<td>2009/10</td>
<td>2009/10</td>
</tr>
</tbody>
</table>
The most influential points for each of the impact metrics are 2010/11, 2009/10, and 2009/10, respectively. This result is not surprising since 2009/10 and 2010/11 take on even greater significance in the relatively short time series of rail-impact data, which is unlikely to sample the full range of impact values. Nevertheless, drawing on the findings for all three transport sectors (aviation, road, and rail), there is no evidence to suggest that the impacts in 2009/10 and 2010/11 are abnormal, given their extreme NAO indices. As such, they represent valid, albeit extreme, data points that can inform our understanding of the impacts of weather on U.K. transport. Given the short time series of available rail-impact data, however, it is wise to be more cautious about the precise nature of that relationship than for the other sectors discussed in this paper. Also, there were no significant lag-1 autocorrelations in any of the rail-impact metric time series.

6. Utility of risk-based impacts forecasts

Table 6 summarizes the quality of fits performed for the nine main impact metrics discussed in sections 3, 4, and 5 above and illustrated in Figs. 2, 3, and 4. The final three table entries are for the combined road-impact metrics, shown in Fig. A1 in the appendix. All but one of the nine main correlations between observed NAO and impact are statistically significant at the 95% level. The exception is marginally significant (p value of ~0.06–0.07). For the forecast NAO, only three of the nine main correlations are statistically significant at the 95% level. Four of the “insignificant” correlations are actually marginally significant (p value of ~0.05–0.07), which suggests that with larger sample sizes as many as seven of the nine main correlations could be statistically significant.

In some cases (rail impacts and salt usage—both of which are relatively small datasets) the correlations are strongly influenced by the severe impacts in the winters of 2009/10 and 2010/11. In addition, the use of salt-usage data at annual resolution rather than seasonal resolution may have affected the fit quality.

In most cases studied here, the quality of the regression of impact metric against NAO index is stronger for the relationship with the observed NAO than for the forecast NAO, typically by a difference of ~0.1–0.2 in

<table>
<thead>
<tr>
<th>Impact metric</th>
<th>$R$ for fit against observed NAO index</th>
<th>$R$ for fit against forecast NAO index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cold impacts</td>
<td>Wet/wind impacts</td>
</tr>
<tr>
<td>Percent of serious road accidents in Great Britain in snowy and calm weather (1)</td>
<td>$-0.68$</td>
<td>$0.55$</td>
</tr>
<tr>
<td>Percent of serious road accidents in Great Britain in wet and windy weather (1)</td>
<td>$-0.61$</td>
<td>$-0.31$</td>
</tr>
<tr>
<td>Aircraft deiced at LHR (1)</td>
<td>$-0.65$</td>
<td>$-0.45$</td>
</tr>
<tr>
<td>Highways Agency annual salt usage (1)</td>
<td>$-0.65$</td>
<td>$-0.45$</td>
</tr>
<tr>
<td>Percent of weather-related delays to flights at LHR</td>
<td>$-0.65$</td>
<td>$-0.67$</td>
</tr>
<tr>
<td>Length of weather-related delays at LHR</td>
<td>$-0.61$</td>
<td>$-0.64$</td>
</tr>
<tr>
<td>Incidents on the railway network in Great Britain for infrastructure categories (Table 4 except 701D MW)</td>
<td>$-0.64$</td>
<td>$-0.64$</td>
</tr>
<tr>
<td>Incidents on the railway network in Great Britain for T&amp;RS category (Table 4 701D MW only)</td>
<td>$-0.64$</td>
<td>$-0.64$</td>
</tr>
<tr>
<td>Incidents on the railway network in Great Britain for all categories</td>
<td>$-0.64$</td>
<td>$-0.64$</td>
</tr>
<tr>
<td>Percent of serious road accidents in Great Britain in snowy and (calm or windy) weather</td>
<td>$0.25$</td>
<td>$0.25$</td>
</tr>
<tr>
<td>Percent of serious road accidents in Great Britain in windy and (fine or wet or snowy) weather</td>
<td>$0.53$</td>
<td>$0.53$</td>
</tr>
</tbody>
</table>
the correlation value \( R \). Despite the small sample size in some cases, this result is consistent with the known performance of the forecast model, which does not perfectly reproduce the observed NAO. As such, it gives further confidence in the validity of our results for NAO–impact relationship. The significance tests that we have performed do not allow for autocorrelation of the NAO from year to year (this autocorrelation is \( 0.1 \)), but correcting the correlation for this effect results in only a very small change in \( R (0.01) \).

It is important to note that in many of the cases investigated here the incidents/impacts that can be predicted by use of the NAO index tend to be a small fraction of the overall set of incidents/impacts. Although this is the case, of course, these predictable elements are frequently among those that cause the highest-profile effects on the transport networks. We have not explored the “onward effects” of the impact metrics used in this paper; for example, we have used incident-count data for rail-related incidents but have not assessed the extent to which those incidents resulted in (say) financial penalties or reputational impact.

To highlight further the potential value of the forecasts for operational planning, we have summarized the findings with contingency tables. We have determined the average impact and then assessed how often the forecast NAO index correctly predicts above- or below-average impacts. For definitions of the hit rate (HR) and false-alarm rate (FAR) see appendix B. The results are given in Table 7; in all cases (except the occurrence of road accidents in wet weather), the HR exceeds the FAR, implying skill for predicting transport impacts. The most successful cases appear to be those for rail impacts, BA weather-related
delays, and salt usage by the Highways Agency, although this has to be considered in conjunction with the correlation significance information in Table 6.

We also calculated the Heidke skill score (HSS; Heidke 1926), using the $2 \times 2$ contingency tables and the following definitions: “event observed” = above-average impacts occurred and “event forecast” = negative NAO was forecast (if linear impact–NAO relationship is negative) or positive NAO was forecast (if linear impact–NAO relationship is positive). HSS measures the improvement of a forecast with respect to chance and may take values from $-\infty$ to 1. A perfect forecast has HSS = 1, a forecast with no skill has HSS = 0, and a forecast that is worse than chance has HSS < 0. The HSS values are given in Table 7; HSS is positive in all cases except one, lending further support to the potential utility of these forecasts.

An illustrative way of quantifying the association between categorical variables in a contingency table is to calculate the odds ratio. For a $2 \times 2$ contingency table of the form

$$
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}
$$

the sample odds ratio can be written

$$
\text{OR} = \frac{a \times d}{b \times c}.
$$

(1)

For the contingency tables above, an odds ratio of 1 indicates that there is no association between NAO index and impact. An odds ratio > 1 indicates a positive association, and a value of < 1 indicates that the association is negative. For moderate to large sample sizes (i.e., greater than $\sim 30$), the sample odds ratio is a reasonable estimator of the population odds ratio. For the small sample sizes that are considered here (7–20 data points), sampling effects should be considered, and it is necessary to have a method of assessing whether an apparent association is nonrandom. To do this, it is standard to use Fisher exact tests (Sprent 1993). These tests tend to be used for small sample sizes but are valid for all sample sizes; in the limit of large sample sizes, $\chi^2$ tests may also be used.

The null hypothesis for the Fisher exact test is that there is no association between the categorical variables. To test for the presence of associations in the contingency tables shown above, we performed one-sided tests. These showed that the null hypothesis could only be rejected at the 90% confidence level for the “percentage of BA LHR delays due to weather” ($p$ value = 0.055), suggesting the existence of an association in this case. For all other impact metrics, the null hypothesis could not be rejected. Given the considerable uncertainty in the sample odds ratios, however, we consider many of these nonrejections to be marginal. For example, the $p$ value of 0.12 for the length of weather-related delays to BA LHR flights is small enough to suggest that a larger sample of impact data is needed to draw more robust conclusions about the association. The same argument applies to road accidents in windy weather ($p$ value = 0.16), rail impacts: infrastructure ($p$ value = 0.18), rail impacts: infrastructure plus T&RS ($p$ value = 0.18), and road-salt usage ($p$ value = 0.23). For the remaining impacts, the $p$ values are $> 0.3$, suggesting that it is unlikely that associations would become statistically significant as a result of including additional data.

This analysis is based on the available impact and forecast data, which amount to, at most, 20 data points being sorted into four categories. The small amount of available data means that the odds ratio can change appreciably by adding or removing a few points. In addition, given that the correlations presented in sections 3, 4, and 5 are inherently noisy, we would expect a fraction of points to fall into the “false alarm” and “miss” categories. Given the small number of data points, it would only take a few points in these categories to mask an association, that is, to produce a $p$ value $> 0.1$ for the odds ratio. As such, it is difficult to draw firm conclusions about the apparent lack of statistically significant associations, except that more data is needed. It is reassuring that, given the sensitivity described above, the odds ratio perhaps represents a more stringent test of the existence of an association than does a standard correlation analysis.

Further support for the suggested skill in these forecasts can be obtained by considering the data in Table 7. Despite it being difficult to ascertain significance in the individual cases shown in the table, the HR exceeds the FAR in 11 of 12 cases. We can use the binomial distribution to calculate the probability of obtaining this result by chance, under the following assumptions: 1) the trials are independent—that is, the same impact data are not used in more than one contingency table—and 2) for each impact metric, the HR and FAR are independent, uniformly distributed random variables. To ensure that the first assumption is satisfied, we omit “rail impacts: infrastructure plus T&RS” from the calculation, since the impact data are used separately in the infrastructure and T&RS assessments. Likewise, the three categories “road accidents in wet weather,” “road accidents in snowy weather,” and “road accidents in windy weather” are omitted because they combine other data types (see explanation in section 4). From the second assumption, it follows directly that probability $P(\text{HR} \geq \text{FAR}) = 0.5$. Using this information, the probability of the HR exceeding
the FAR in 7 or more of 8 cases by chance is $\sim 0.035$, which suggests overall skillful prediction of the chances of transport-related impacts.

7. Discussion

**Links between meteorological conditions and impacts, and the potential for impacts forecasts on seasonal time scales**

The essential requirement being addressed in this paper is shown schematically in Fig. 5. In both forecast and observed space, an implicit relationship exists between a meteorological driver—in this case, the NAO index—and some “derived quantity”—in this case, the number of extreme events brought about by the meteorological driver (e.g., snow days or very wet days—links 1 and 4 in Fig. 5). In turn, there is a relationship between these extreme events and their impact on the transport system (links 2 and 5 in Fig. 5).

The ultimate goal is to find some skillful method for forecasting observed impacts, which is represented conceptually by the thick dashed arrow in Fig. 5. One way to do this would be to make use of all the relationships (links 2, 1, 3, 4, and 5, in that order), for example, by making use of a coupled impacts model driven by the meteorological forecasts to predict the risk of disruptive events. The degree of noise involved in the relationships between meteorological conditions, derived quantity, and impact can be large, however. For example, the general circulation models used in producing seasonal forecasts may not necessarily represent well the relationship in link 1; similarly, the occurrence of observed snow days or very wet days may not be related solely to the NAO index (link 4) but could also depend on chaotic fluctuations or other circulation patterns (e.g., Fereday et al. 2008). Particular extreme events need not always result in an impact on the transport system (link 5). As such, there is a relatively large number of “noisy” steps to go through by using such a method.

The predictability of the NAO in the seasonal-forecast system used here (Fig. 5, link 3), which has recently been demonstrated (Scaife et al. 2014) to have a correlation of $R \approx 0.62$, means that a forecast system can be developed that in principle involves exploiting only links 3, 4, and 5 in Fig. 5 and eliminating links 1 and 2. In fact, even if we were to take this approach with a perfect forecast system with a coupled impacts model, our approach of using the NAO as a proxy predictor should still be more skillful, with the skill of the perfect coupled impacts–forecast system asymptotically approaching that of our method, for perfect models and large ensembles of forecasts.

By way of illustration, consider the case of developing a forecast using the data for length of weather-related delays at LHR. Here the actual extreme events are the observed number of frost days per winter and the forecast extreme events are the GloSea5 gridpoint frosts. Figure 6 shows correlations between the relevant quantities. The correlation between forecast extremes and observed impacts is not significant, whereas that between forecast NAO and observed impacts is larger and significant. This result supports the above assertion that building an impacts forecast that is based on the forecast NAO index is a more skillful approach than building a forecast that is based on model representation of extremes or impacts.
As illustrated in Figs. 2, 3, and 4 there are several unusual winters that appear to influence both the gradient and strength of correlation between the NAO index and the impact. The most obvious of these is 2009/10, which has the highest leverage in all impact categories. This is not surprising given that 2009/10 exhibited the most negative average winter NAO index on record and given the relatively short time series of available data. Evidence from this investigation does indicate that the impacts in the winter of 2009/10 are broadly consistent with what would be expected from a linear relationship and for such an extreme NAO index. Indeed, given the extremity of the winter 2009/10 NAO index, we may expect it to be more influential than the statistical analysis suggests; for example, according to the statistics of Cook’s distance measure presented in Tables 2, 3, and 5, 2006/07 is the most influential year in the same number of cases as 2009/10 is. We have also seen that 1989/90 is a clear outlier in the relationship between road accidents in wet and windy weather and the NAO as a result of the year experiencing an impact that was significantly greater than would be predicted from the NAO index. This is, perhaps, an indication of the impact of the severe Burns’s Day storm that hit the United Kingdom in January of 1990 (Rowe 1990). In a more general context, it also suggests that subseasonal variations in the NAO index, along with the occurrence of unrelated weather patterns, could be the source of significant scatter in the correlations presented in Figs. 2, 3, and 4 (e.g., Kolstad et al. 2010). Understanding the origin of this scatter will be an important part of future work, since it will enable us to characterize better the nature of the relationship between weather and its impacts on the U.K. transport network. This could be achieved using a more detailed weather-pattern analysis—for example, using clustering methods (Fereday et al. 2008).

8. Summary and conclusions

We have assessed the degree to which it is possible to build useful relationships between the observed and forecast NAO index and a selection of winter impacts on U.K. transport. We have demonstrated that there is potential for such relationships to provide skillful information to transport stakeholders, at time scales well beyond traditional weather forecasts. In particular, we found that, where a priori knowledge exists regarding the likely relationship between NAO and impact (i.e., where the impact metric referred to either cold/snow-related or rain/wind-related impacts), all correlations with both the observed and forecast NAO exhibited the correct sign.

This assessment has been preliminary in nature and could be extended to cover other regions where there is skill in the NAO forecast, such as North America or other parts of Europe. Appropriate impacts data are needed to do this over a long enough record to make used of multi-year retrospective forecasts. Another possible area of further work would be to explore analogous relationships for other transport subsectors—such as shipping, for which recent studies of seasonal predictability of North Atlantic Ocean and North Sea winds and wave heights have shown promise (Colman et al. 2011; Brands 2014). In addition, skill has been demonstrated in seasonal predictions of Baltic Sea ice cover (Karpechko et al. 2015), which could enable better preparation for ice-breaking activities supporting Baltic Sea navigation during winter.

Forward look: Developing real-time forecasts of winter transport disruption

The assessment of the relationships between NAO index and impact metrics discussed above is the first stage in the development of trial real-time forecasts of
TABLE A1. Railway delay categories deemed by Network Rail to be weather related. Here, RHC is railhead conditioning (train), DMU is diesel multiple unit, HST is high-speed train, MPV is multipurpose vehicle, and EMU is electric multiple unit. Here, “NZ Pumps T” and “NZ Pumps F” are obsolete codes covering aspects of autumn delay attribution, as agreed between Network Rail and train operators (T) and Network Rail and freight operators (F).

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Incident reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points failures</td>
<td>101 JT</td>
<td>Points failure caused by snow or frost where heaters are not fitted</td>
</tr>
<tr>
<td></td>
<td>101 IP</td>
<td>Points failure due to snow/frost where heaters are fitted but are not operative or are defective</td>
</tr>
<tr>
<td>Severe weather (beyond design capability of infrastructure)</td>
<td>110A X3</td>
<td>Lightning strike: Damage to protected systems</td>
</tr>
<tr>
<td></td>
<td>110A WX</td>
<td>Miscellaneous items on line (including trees) due to effects of weather; responsibility of IMC</td>
</tr>
<tr>
<td></td>
<td>110A XX</td>
<td>Miscellaneous items on line (including trees) due to effects of weather; responsibility of RT</td>
</tr>
<tr>
<td></td>
<td>110A X9</td>
<td>Points failure caused by severe snow where heaters are working as designed</td>
</tr>
<tr>
<td></td>
<td>110A X2</td>
<td>Severe flooding beyond that which could be mitigated on Network Rail infrastructure</td>
</tr>
<tr>
<td></td>
<td>110A XH</td>
<td>Severe heat affecting infrastructure; responsibility of Network Rail (excluding heat speeds)</td>
</tr>
<tr>
<td></td>
<td>110A WT</td>
<td>Severe snow affecting infrastructure; responsibility of IMC</td>
</tr>
<tr>
<td></td>
<td>110A XT</td>
<td>Severe snow affecting infrastructure; responsibility of Network Rail</td>
</tr>
<tr>
<td></td>
<td>110A XW</td>
<td>Severe weather other than snow affecting infrastructure; responsibility of Network Rail</td>
</tr>
<tr>
<td></td>
<td>110A WW</td>
<td>Severe weather other than snow affecting infrastructure; responsibility of IMC</td>
</tr>
<tr>
<td>Other weather (impact on infrastructure or network operations)</td>
<td>110B OF</td>
<td>Blanket speed restriction for extreme heat or high wind</td>
</tr>
<tr>
<td></td>
<td>110B X4</td>
<td>Blanket speed restriction for extreme heat or high wind in accordance with the Group Standards</td>
</tr>
<tr>
<td></td>
<td>110B JH</td>
<td>Critical rail temperature speeds (other than buckled rails)</td>
</tr>
<tr>
<td></td>
<td>110B JK</td>
<td>Flooding not due to exceptional weather</td>
</tr>
<tr>
<td></td>
<td>110B OG</td>
<td>Ice on conductor rail/OLE</td>
</tr>
<tr>
<td></td>
<td>110B J6</td>
<td>Lightning strike against unprotected assets</td>
</tr>
<tr>
<td></td>
<td>110B IW</td>
<td>Nonsevere: Snow/ice/frost affecting infrastructure equipment</td>
</tr>
<tr>
<td></td>
<td>110B OK</td>
<td>Special working for fog and falling-snow conditions</td>
</tr>
<tr>
<td></td>
<td>110B X1</td>
<td>Visibility in semaphore-signaled areas or special workings for fog and falling snow implemented by Network Rail—in all signaling areas</td>
</tr>
<tr>
<td>Wheel slip due to leaf fall</td>
<td>111A QH</td>
<td>Adhesion problems due to leaf contamination</td>
</tr>
<tr>
<td></td>
<td>111A QI</td>
<td>Cautioning due to railhead leaf contamination</td>
</tr>
<tr>
<td></td>
<td>111A OE</td>
<td>Failure to lay Sandite or operate RHC as programmed</td>
</tr>
<tr>
<td>Low adhesion including autumn (Network Rail)</td>
<td>150 Q2</td>
<td>NZ Pumps F</td>
</tr>
<tr>
<td></td>
<td>150 Q3</td>
<td>NZ Pumps T</td>
</tr>
<tr>
<td>Track circuit failures—leaf fall</td>
<td>305 QJ</td>
<td>Special working for leaf-fall track circuit operation</td>
</tr>
<tr>
<td></td>
<td>305 ZC</td>
<td>Track circuit failures due to leaves</td>
</tr>
<tr>
<td>Network Rail operations—RHC</td>
<td>501 C OS</td>
<td>Delays to other trains caused by RHC taking long time</td>
</tr>
<tr>
<td></td>
<td>501 C OO</td>
<td>Late start of an RHC</td>
</tr>
<tr>
<td></td>
<td>501 C OM</td>
<td>Technical failure associated with an RHC</td>
</tr>
<tr>
<td>Technical fleet delays</td>
<td>701 D MC</td>
<td>Diesel locomotive failure/defect/attention: Traction</td>
</tr>
<tr>
<td></td>
<td>701 D MD</td>
<td>DMU (including HST)/MPV failure/defect/attention: Traction</td>
</tr>
<tr>
<td></td>
<td>701 D MB</td>
<td>Electric locomotive (including IC225) failure/defect/attention: Traction</td>
</tr>
<tr>
<td></td>
<td>701 D MM</td>
<td>EMU failure/defect/attention: Traction</td>
</tr>
<tr>
<td></td>
<td>701 D MW</td>
<td>Weather: Effect on T&amp;RS equipment</td>
</tr>
<tr>
<td>Low adhesion including autumn (train operator)</td>
<td>750 FT</td>
<td>Leaf-fall neutral</td>
</tr>
<tr>
<td></td>
<td>750 TT</td>
<td>Leaf-fall neutral</td>
</tr>
<tr>
<td></td>
<td>750 MP</td>
<td>Locomotive/unit adhesion problems</td>
</tr>
</tbody>
</table>
winter transport disruption on the U.K. transport system, through a project coordinated by the U.K. government Department for Transport and involving stakeholders across the U.K. transport system, including Network Rail, British Airways, the Highways Agency, the Association of Train Operating Companies, the British Ports Association, and others.

In cases in which a significantly skillful relationship is found between NAO index and impact metric, this information could in principle be used in conjunction with real-time seasonal forecasts to develop a risk-based forecast for particular types of winter transport disruption in the United Kingdom, ahead of each winter. A further consideration in the development of any forecast is how it could influence the decision(s) made by stakeholders; this would require clarification through stakeholder liaison and more detailed probabilistic assessment than that presented here. Therefore, further work will seek to develop and try out these impact-based forecasts through the European Union’s Seventh Framework Programme (EU FP7) European Provision of Regional Impacts Assessments on Seasonal and Decadal Time Scales (EUPORIAS) project, whose goal is to deliver prototype climate services at the seasonal-to-decadal time scale. An example of how the forecasts could be used is to inform decisions about replenishment of stocks of deicing materials (e.g., fluid for aircraft and salt for roads). Preliminary investigation with stakeholders has suggested that—although decisions about initial stock levels ahead of winter are sometimes taken further in advance than the lead times offered by these forecasts—the information could be useful within a given winter, in

FIG. A1. Relationship between either (left) observed or (right) forecast NAO index and percentage of serious road accidents in winter (DJF) occurring during (top) windy, (middle) wet, and (bottom) snowy conditions. For each panel, the straight line indicates the linear regression fit; dotted curves give the 95% confidence intervals on the fit. Correlation coefficients and p values on the fits are specified at the top right of each panel.

<table>
<thead>
<tr>
<th>TABLE B1. Generalized contingency-table form used in Table 7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ NAO</td>
</tr>
<tr>
<td>Above-average impact</td>
</tr>
<tr>
<td>Below-average impact</td>
</tr>
</tbody>
</table>
conjunction with other forecasts, when making decisions such as whether to procure more deicing material, how much to procure, and when to run down stocks toward the end of winter.

The theoretical limit in predictability of the NAO is limited by forecast ensemble size and is therefore likely higher than the $R \sim 0.6$ in the hindcast set used here. It will be a number of years before a formal assessment can be made of the utility of impact forecasts of this nature, but, along with more careful development of the weather–impact relationships, it appears that skillful seasonal forecasts of transport impacts are now becoming possible.

Acknowledgments. This work was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreements 308291 and 308378 (EUPORIAS and SPECS projects), and from the U.K. government Department for Transport. We gratefully acknowledge the Science and Research Team (U.K. government Department for Transport) for support and coordination of the transport stakeholder group and the organizations that provided industry data to support our analysis: Network Rail data were provided by Lesley Potter (Network Rail Performance Analysis team), British Airways data were provided by Peter Lynam and Michael Ward, Highways Agency data were provided by Robin Herringshaw and Paul Furlong, and STATS19 data were obtained from the Administrative Data Liaison Service (http://adls.ac.uk). We thank Kate Brown for useful discussions on statistical testing and metrics.

APPENDIX A

Further Information about Rail and Road Impacts

Table A1 shows the full range of incident causes deemed by Network Rail to be weather related. A subset of these was used in this analysis (see Table 4).

Figure A1 shows the correlation results for the combined accident causes that are discussed in section 4.

APPENDIX B

Hit Rate and False-Alarm Rate

In Table 7 we have defined $2 \times 2$ contingency tables of the form shown in Table B1. For cases in which the correlation between NAO and impact is negative, a negative NAO index is expected to result in above-average impacts. A forecast hit then involves a combination of forecast negative NAO index and above-average impacts, and therefore the contingency table is populated as in Table B2, where $F = \text{false}$, $T = \text{true}$, $N = \text{negative}$, and here $P = \text{positive}$. The HR (sometimes also called the probability of detection) and the FAR (sometimes also called the probability of false detection) are then defined as

$$HR(\%) = 100\% \times \frac{\text{hits}}{\text{hits} + \text{misses}}$$
$$= 100\% \times \frac{b}{a + b}$$  and

$$FAR(\%) = 100\% \times \frac{\text{false alarms}}{\text{false alarms} + \text{correct rejections}}$$
$$= 100\% \times \frac{d}{c + d}$$

For cases in which the correlation between NAO and impact is positive, a positive NAO index is expected to result in above-average impacts. A forecast hit then involves a combination of forecast positive NAO index and above-average impacts, and therefore the contingency table is populated as in Table B3, and consequently

$$HR(\%) = 100\% \times \frac{a}{a + b}$$  and

$$FAR(\%) = 100\% \times \frac{c}{c + d}.$$

The two different cases have been accounted for in the calculation of HR and FAR in Table 7.

REFERENCES


Unauthenticated | Downloaded 07/28/22 09:47 AM UTC


