The Journal of Risk Finance

# Quantifying spatial basis risk for weather index insurance

Michael T. Norton, Calum Turvey, Daniel Osgood,

Received July 2012
Revised September 2012
Accepted September 2012

# Quantifying spatial basis risk for weather index insurance

Michael T. Norton
*International Research Institute of Climate and Society (IRI), Columbia University, Palisades, New York, USA and University of California, Davis, Davis, California, USA*

Calum Turvey
*Dyson School of Applied Economics and Management, Cornell University, Ithaca, New York, USA, and*

Daniel Osgood
*International Research Institute of Climate and Society (IRI), Columbia University, Palisades, New York, USA*



## Abstract

**Purpose** – The purpose of this paper to develop an empirical methodology for managing spatial basis risk in weather index insurance by studying the fundamental causes for differences in weather risk between distributed locations.

**Design/methodology/approach** – The paper systematically compares insurance payouts at nearby locations based on differences in geographical characteristics. The geographic characteristics include distance between stations and differences in altitude, latitude, and longitude.

**Findings** – Geographic differences are poor predictors of payouts. The strongest predictor of payout at a given location is payout at nearby location. However, altitude has a persistent effect on heat risk and distance between stations increases payout discrepancies for precipitation risk.

**Practical implications** – Given that payouts in a given area are highly correlated, it may be possible to insure multiple weather stations in a single contract as a "risk portfolio" for any one location.

**Originality/value** – Spatial basis risk is a fundamental problem of index insurance and yet is still largely unexplored in the literature.

**Keywords** Insurance, Precipitation, Rainfall, Index insurance, Weather derivatives, Spatial basis risk, Basis risk

**Paper type** Research paper

## Introduction

Basis risk, the difference between the insured quantity and the underlying risk, is known to be a major concern with weather index insurance. Because weather index insurance insures an "index" of weather variables instead of actual adjusted losses, the settlement of the index insurance may not match the risk exposure of the insuree. Basis risk may

**JEL classification** – C40

The authors would like to thank contributors to the Myers endowment fund and the Risk Management Agency Partnership Agreement USDA-02IE08310229. The authors would also like to thank conference participants at the AAEA Annual Meeting in Denver, Colorado, in July 2010 and the NAREA workshop in Atlantic City, New Jersey, in June 2010.



1526-5943
DOI 10.1108/15265941311288086

take a spatial dimension, as the coverage of weather stations across the world is not perfect, or a local dimension, for example, if the index/yield relationship were found to be flawed.

This paper investigates the problem of spatial basis risk for weather index insurance, which is necessarily a complicated subject of investigation. Part of the complication is that pricing weather index insurance (or weather derivatives) according to burn rate analysis includes not only a spatial dimension but also a temporal one as historical frequencies are calculated. Weather risk is further subjected to potential long-term trends due to climate change as well as variations due to prevailing weather conditions such as the El Niño Southern Oscillation (ENSO) index.

From an empirical point of view, this paper investigates basis risk using a decorrelation approach to space measurement. We have developed a program that is linked to all of the weather stations in a given region (i.e. includes data on all weather stations in the USA). From a randomly selected point, we select all weather stations within a certain radius and calculate the particular weather risk for each year at each station to calculate the burn rate insurance premium. This is done for the same specific criteria for all stations within the sphere. Next, we compare on a year-by-year basis the payouts that would have been made at each pair of locations. Thus, if there are $n$ weather stations then there are $n(n-1)/2$ pairwise comparisons. Finally, we measure the error or basis risk between each station pair and regress the mean differences against a number of spatial variables. These spatial variables include longitude and latitude coordinates as well as elevation difference and distance between weather stations.

This exercise reveals the spatial characteristics which have a persistent effect on temperature and precipitation risk. For temperature/heat risk, the most important independent variable is the difference in elevation between stations, explained by the fact that the primary temperature difference will be due to the elevation of the station. Rainfall correlation, however, is strongly dependent on distance as precipitation will often be an extremely localized phenomenon, with the other variables of lesser importance.

Lastly, we offer a pricing strategy for both temperature and precipitation risk tailored to the specific typology of each distinct type of risk. In many cases, the "portfolio" method of selecting a proportion of risk from each nearby station based on similarity in elevation (temperature) or geographic placement (precipitation) offers advantages over an index based more sophisticated spatial statistics algorithms. By compartmentalizing the risk into existing weather stations, the "portfolio" method has the advantage of allowing for easier pricing of policies for insurance and reinsurance companies. In this manner, we will demonstrate a strategy for pricing risk in distributed locations around the USA.

This paper is structured as follows: first, a discussion of the mathematical and spatial considerations for analyzing weather risk. This is followed by the introduction of a regression equation that attempts to predict differences in risk at existing stations through the use of the geographic characteristics of those stations. The results of the regression equation lead to a discussion of a pricing strategy for both temperature and precipitation risk tailored to the specific typology of each distinct type of risk as revealed by the regression results.

## Background

Weather index insurance is a recent financial innovation that has received much attention from academics and implementers alike as a way to smooth risk for agricultural producers (Zeng, 2000; Turvey, 2001; Vedenov and Barnett, 2004). By using an “index” of weather observations as a proxy for crop loss, the problems of traditional indemnity insurance are reduced or eliminated. Weather index insurance removes the subjective nature of insurance adjustment as well as the problems of adverse selection and moral hazard that are present in the traditional indemnity insurance model. Weather index insurance makes it possible to offer microinsurance to rural farmers in developing countries, which can serve a valuable function in a development intervention and may lead to more interactive benefits, such as improved access to rural credit (Skees, 2008). In countries such as China (Turvey and Kong, 2010; Göncü, 2011), India (Seth *et al.*, 2009), and Portugal (Ghiulnara and Viegas, 2010), among others, researchers have studied the feasibility of weather index insurance as a risk hedging mechanism.

However, despite the promise of the technology, it is not always straightforward to apply. In particular, we trade the problems of adverse selection and moral hazard with that of basis risk, which is defined as the risk that payoffs of a hedging instrument do not correspond to the underlying exposures. Basis risk may be reduced through the selection of appropriate weather observations to construct the index, but in reality the prevailing weather conditions are only one variable in crop production and are often considered exogenous to the production function (Turvey and Norton, 2008). Basis risk is a major problem when using a risk-smoothing implement such as weather index insurance. The good years should help pay for the bad years, but if the product were not aligned properly, this strategy of risk smoothing could be harmful to the producers’ bottom lines. Basis risk as a problem for index insurance is widely acknowledged in the literature and has even been seen to affect take-up rates in India (Giné *et al.*, 2008).

The proposed solutions to the problem of basis risk are varied. Heimfarth and Musshoff (2011) quantify spatial basis risk in the North China Plain using a decorrelation function. One approach is to perform spatial analysis techniques on weather data to provide a historical time series in varied geographic locations (Paulson *et al.*, 2010). Another study intentionally analyzed data from a flat area with consistent elevation (Richards *et al.*, 2004). Other researchers link microinsurance to microcredit and advocate for a central financial institution to aggregate index insurance contracts so as to average out basis risk for all actors (Miranda and Gonzalez-Vega, 2011; Woodard and Garcia, 2008b). If index insurance is to be widely used as a risk mitigation and climate adaptation tool for individual farmers, the problem of basis risk must be better understood.

A traditional difficulty in pricing weather index insurance is that there are only a certain number of weather stations for which historical data exists. A longer time series of data provides more confidence for historical burn rate analysis pricing and a minimum number of years is needed to understand historical weather patterns, which will often require at least ten or 20 years of weather data. Establishing a new station can provide high quality weather data, but there will be no historical record at that new station with which to price risk. This problem is acute in countries with poor infrastructure, which is paradoxically where weather index insurance might do the most good (Morduch, 2006). But even in places with many long-established weather stations, the spatial distribution of risk is not yet fully understood. The challenge that is present in the weather index insurance market is how to strategically leverage the information from existing stations at geographic locations where the precise weather observations are unknown.

**Theoretical basis**

Some researchers have applied spatial analysis techniques to the weather observations directly, and used that information to construct a surface of historical time series observations for any geographic point. While techniques like kriging have shown to be very accurate in producing a prediction surface for points in space, these spatial analysis techniques are not designed to model for deviations from normal conditions, which is precisely what we are interested in and want to protect against.

This paper takes rather the opposite approach to spatial analysis by pricing risk at known locations and analyzing how that risk changes through space. To that effect, we have developed a web-based computer program, as described in Turvey and Norton (2008), that is able to analyze historical weather observation data for all weather stations in the USA. From a randomly selected point we select all weather stations within a certain radius and calculate the particular weather risk at each station to calculate the burn rate insurance premium.

**Mathematical considerations for spatial weather risk**

Table I lists aggregate temperature and rainfall observations for Ithaca, NY for June 1-August 31 along with the mean observation for all stations within a proscribed radius (100 miles for temperatures and 67 miles for rainfall). The overall means are similar, but when we examine the yearly variation as measured by the average correlation between the base station (Ithaca) and every other station, we find that heat is highly correlated but rainfall less so. A familiar pattern is that when we introduce risk events (defined later), the variability increases, not only in the averages but also in the correlation. Our risk events were intentionally defined as generally as possible and represent a kind of baseline for analyzing weather risk. One way that we may think of this is that the weather observations themselves have a certain structure and correlation, but we attempt to impose order upon those observations in the form of risk events, we lose some of the structure of that information. These risk events also represent relatively common events over long date ranges; presumably these numbers would also weaken if a more specific time frame or risk event were used.

The challenge presented is, very simply, to improve the accuracy of the yearly correlation. Although this may seem somewhat abstract, insurance policies have profound real-world implications for farmers holding a policy and it is crucial to match the years with payments with the actual losses. By taking the payout schedule for all stations and adjusting for geographic variables, we can potentially price insurance contracts for any given point on the map. Because of the vast number of stations located around the country, our hopeful result is a simple equation in which we can build upon this simple methodology and adjust for the differences in distance, altitude,

| Weather risk | Base station | Mean of surrounding stations | Difference (%) | Correlation |
|---|---|---|---|---|
| Heat risk event (payout) | \$22.37 | \$27.31 | −22.10 | 0.78 |
| CDD index (85°F) | 68.41 | 71.51 | −4.50 | 0.89 |
| Drought risk event (payout) | \$20.95 | \$22.61 | −7.90 | 0.69 |
| Cumulative rainfall (in) | 10.65 | 10.36 | 2.60 | 0.75 |

**Table I.** Correlation of average of nearby stations of cumulative weather indexes and polar coordinates. What follows is an attempt to provide a universal solution using those readily available geographic variables to arrive at a payout for any unknown location.

**Defining the risk events**

Choosing an event that is sufficiently general yet meaningful for all sites is difficult, because there is no such thing as generality. For example, a heat event in upstate New York is incomparable to a heat event in a warmer climate. Temperatures in central New York infrequently reach above 90°F, but in Norman, OK this temperature is reached quite frequently in summer months (Turvey and Norton, 2008). The sheer variation of climates in America requires us to tailor our heat risk events for each station.

To start, evidence indicates that temperatures above 85°F correlate with crop yield losses (Schlenker and Roberts, 2006). Using this as a benchmark, we employ a common financial tool to measure long-term deviations in temperature observations known as degree days. Well established in the derivatives market, degree days accumulate above or below a benchmark value and indicate the relative suitability of a given time period. Because only degrees of temperature that are above the benchmark value are accumulated, a degree day measure is quite appropriate in this instance as we are modeling crop losses due to excessively high temperatures.

For this paper, we use a cooling degree day (CDD) index with benchmark value of 85°F and the mean CDD at the base station serving as the strike value (or "trigger") with a sliding payout for values above that. Payouts are calculated at each station for every year data is available. Figure 1 shows the payout schedule for Ithaca, NY, where mean CDD is 68.41.

Six weather stations were selected at sites across the country according to quality of data and the absence of geographic features within 50 miles that would prevent weather stations from being placed, such as bodies of water or international borders. Mean CDD for those six stations varies from 68.41 at Ithaca, NY to 720.63 in Davis, CA, and are listed in Table II.

For precipitation, the contract is identical for all sites. We use a drought event of less than 0.1″ of precipitation over any 14-day period. The payout occurs on a sliding scale

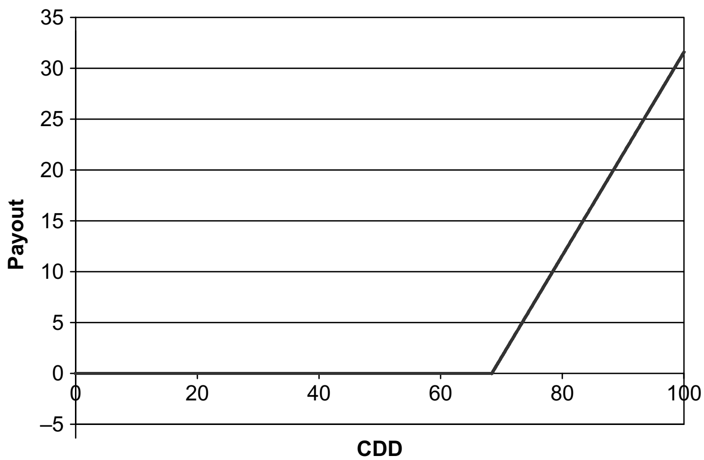


**Figure 1.**
Schedule of payouts for heat risk event

with \$10 accumulating for each hundredth of an inch less than 0.1″, to a maximum of \$100 per event if no rainfall was recorded. Up to three non-overlapping events are possible, with an annual maximum liability of \$300. Figure 2 shows the payoff schedule due to the observed rainfall in any 14-day period, but yearly payoff amounts range from \$0 to \$300 because of the possibility of multiple events.

### Defining the geographic area

Weather Wizard is flexible as to the distance of the radius extending from the base station, but there are a few requirements that must be considered for a successful trial. A certain number of stations are needed to provide contrast, but there are relatively few stations within a short distance (ten miles) of each other. However, as we increase the radius of the circle, the area of the circle increases exponentially. Barring any obstacles like oceans or international borders, the number of stations increases exponentially as the radius of the circle increases. Because we compare each station against each other in each year, this also dramatically increases the number of comparisons that are made, as given by the following formula:

$$comparisions = \frac{n*(n-1)}{2}*years$$

where $n$ is the number of stations within the selected geographic radius, and years is the number of years of data at the base station. The total number of comparisons is subject to missing and incomplete data; many stations have only limited data, and with longer time horizons the potential for periods of missing data within years becomes greater.

| Station | Bridgeport, NE | Bethany, MO | Greenville, AL | Davis, CA | Ithaca, NY | Mosquero, NM |
|---|---|---|---|---|---|---|
| Mean CDD (85°F) | 442.13 | 350.74 | 603.18 | 720.63 | 68.41 | 282.05 |

**Table II.**
Mean CDD (benchmark 85°F) at each location

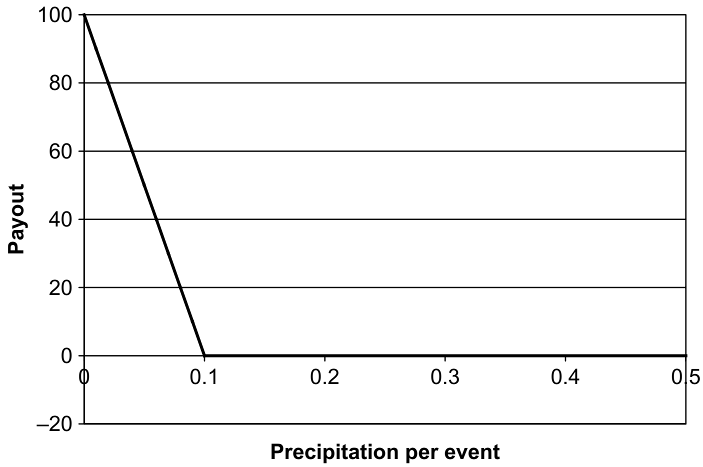


**Figure 2.**
Schedule of payouts for drought risk event

Table III displays the number of stations for each type of weather data within a certain number of miles. Using the number of stations as the value of $n$ in the equation above as well as the number of years of data at the base station, the number of potential comparisons between years is calculated.

Such factors as length of contract and number of years selected will also affect the percentage of available data as presented in Table III. In particular, these percentages are somewhat low because of a relatively long date range. In this case, a 92-day window encompassing June-August was selected, which offers more opportunities for data to be missing than a more carefully targeted risk event. Also, more importantly, very few stations have data continuously to 1,926 as Ithaca does; most stations date to just after Second World War, and it is not uncommon for a station to have as little as one or two years of data for the entire 75 year period. If we selected a shorter contract length (say, 15 days instead of 92) fewer stations would be disqualified for missing data; likewise, if we only considered years after 1949, the percentage of actual comparisons would improve markedly. This discussion is intended to underscore the fact that even though we might define an identical geographic area, there is often a very different spatial distribution of data within that area depending on the parameters we select.

Also, perhaps in acknowledgement of the periodic, unpredictable nature of rainfall, precipitation observation stations are more densely placed and often contain more years of data. In the case presented here, there are more than twice as many precipitation gauges in a given radius than temperature stations, even if the percentage of usable data is roughly similar. It is very likely that temperature observations are placed more sparsely to reflect the intuition that temperatures are considered to vary more continuously over a geographic area, while rainfall can be an extremely localized event.

The advantage of this comparison-based model is that it treats all weather stations equally and is able to include otherwise useless data. In this model, the data will be compared on a year-by-year basis, regardless of how many years of data are at a particular station. The weather stations that only have a few years of data help provide contrast for spatial distributions of risk even though it is impossible to accurately price a contract for that station individually.

| Miles | Stations | Potential comparisons | Actual comparisons | Percentage |
|---|---|---|---|---|
| *Rainfall* | | | | |
| 10 | 2 | 225 | 93 | 41.33 |
| 15 | 4 | 750 | 223 | 29.73 |
| 20 | 12 | 5,850 | 801 | 13.69 |
| 25 | 17 | 11,475 | 1,628 | 14.19 |
| 30 | 25 | 24,375 | 3,304 | 13.55 |
| 35 | 35 | 47,250 | 6,921 | 14.65 |
| *Heat* | | | | |
| 10 | 2 | 225 | 23 | 10.22 |
| 15 | 2 | 225 | 23 | 10.22 |
| 20 | 4 | 750 | 105 | 14.00 |
| 25 | 6 | 1,575 | 220 | 13.97 |
| 30 | 10 | 4,125 | 525 | 12.73 |
| 35 | 16 | 10,200 | 1,366 | 13.39 |

**Table III.**
Number of comparisons in Ithaca, NY for a given number of miles

Also of pertinent interest is what these details entail for selecting a radius to study. As the radius increases, the area of study increases exponentially (according to the area of a circle – $\pi r^2$). The number of stations increases accordingly, which has vast ramifications for the number of potential comparisons according to the equation above. Since Weather Wizard is hosted on a web platform, there are limitations to the amount of data that it can process in a single iteration – selecting a radius requires the user to select a value large enough to offer meaningful results that will also fit within technical possibilities. For this paper, we are using a radius of 50 miles, which is large enough to allow the inclusion of sufficient stations for both heat and precipitation, but small enough to run properly on the Weather Wizard web site.

## The regression equation

The goal when formulating this regression equation was to try and predict the difference in payouts in any given year between any two locations using simple geographic variables:

$$(P_1 - P_2) = \beta_1\varphi + \beta_2(\alpha_1 - \alpha_2) + \beta_3(\omega_1 - \omega_2) + \beta_4(\lambda_1 - \lambda_2) + \beta_0 + \varepsilon$$

where $P_x$ are payouts at station 1 and 2, $\varphi$ is the distance between the two stations, $\alpha_x$ is the altitude at each station, $\omega_x$ is the latitude at each station, and $\lambda_x$ is the absolute value of the longitude of each station (as longitudes in the western hemisphere are traditionally negative).

This equation is primarily a difference equation, where we are attempting to explain the difference in payouts by the difference in altitude and geographic coordinates. At first glance, it seems as if the $\varphi$ variable, distance, is ill-suited for inclusion because distance is strictly positive, and the differences in any part of the equation can easily be negative. However, by imposing a condition of $P_1 \geq P_2$ and adjusting the order of the independent variables to match the condition, we may ensure symmetry between the left and right sides of the equation; only if $(P_1 - P_2)$ is strictly positive will it reflect a potential linear relationship with $\varphi$. Furthermore, distance is a trigonometric function of the individual latitude and longitude variables but is highly correlated to neither. This is because it is a joint function of latitude and longitude, and a degree of longitude is not a constant surface measurement but varies according to distance from the pole. It is useful to think of the latitude/longitude coordinates as reflecting directionality, and distance as an adjustment for increasing variability at increased distances.

The equation for distance is given thusly:

$$\varphi = R{*}Cos^{-1}(Sin(\omega_1){*}Sin(\omega_2) + Cos(\omega_1){*}Cos(\omega_2){*}Cos(\lambda_2 - \lambda_1))$$

where $R$ is a constant reflecting the radius of the sphere we can use to normalize to standard units; the constant for miles is 3,963.1.

What we are left with is a description of how each station compares to each other in three-dimensional space, not only in distance ($\varphi$) but with $x$ and $y$ coordinates given by the latitude ($\omega_x$) and longitude ($\lambda_x$), and z coordinate given by altitude ($\alpha_x$). Our initial hypothesis was that distance ($\varphi$) should be positively correlated in both heat and precipitation, meaning that as distance increases, so do the differences in premiums. For rainfall, the rest of the geographic variables are indeterminate, given that coordinates and/or altitude would seemingly have no effect on the sporadic nature of rainfall. For heat, however, we might expect that altitude and latitude have a negative effect on risk; or, in other words, heat risk is decreased by either an increase in elevation or more northerly locations.

## Regression results

The first thing to notice when looking at these results is that (with one exception) the $R^2$ values are usually quite near to zero, which is to say that these geographic variables provide a very poor explanation for differences in payout amounts between stations. In and of itself this is scant evidence for the predictive power of the geographic variables on the differences in payouts, but most of the difference is accounted for in the constant term. The coefficients for the geographic variables are quite often significant, but unpredictably so. How much of the unpredictability is due to localized conditions is impossible to determine, but local geographic idiosyncracies might have effects on the latitude/longitude coefficients, as directionality within different locations could reflect different geographic characteristics (Tables IV and V).

The two enduring relationships that can be deduced are the effect of altitude on heat risk and the effect of distance on precipitation risk. The signs are consistent and significant for all stations except Greenville, AL. The coefficient for altitude for the heat risk

**Table IV.**
Regression results for heat risk event

| Station | Bridgeport, NE | Bethany, MO | Davis, CA | Ithaca, NY | Mosquero, NM |
|---|---|---|---|---|---|
| No. of years | 104 | 75 | 83 | 74 | 71 |
| Mean CDD | 442.13 | 350.74 | 720.63 | 68.41 | 282.05 |
| Stations within 50 miles | 19 | 25 | 44 | 35 | 16 |
| $n$ | 4,300 | 4,417 | 7,255 | 5,953 | 1,831 |
| $R^2$ | 0.01 | 0.04 | 0.02 | 0.05 | 0.49 |
| Distance | −0.22** | 0.12 | 0.29** | 0.04 | 0.39* |
| Alt. diff. | −0.01** | −0.15** | −0.02** | −0.01** | −0.21** |
| Lat. diff. | 0.43 | 14.90** | 31.28** | −15.93** | 60.89** |
| Long. diff. | −9.29** | 15.13** | 18.27** | 7.85** | 84.17** |
| Constant | 52.30** | 35.23** | 80.51** | 20.72** | 69.95** |

**Note:** Significance at: *10 and **5 per cent levels, respectively

**Table V.**
Regression results for drought risk event

| Station | Bridgeport, NE | Bethany, MO | Davis, CA | Ithaca, NY | Mosquero, NM |
|---|---|---|---|---|---|
| No. of years | 104 | 75 | 83 | 74 | 71 |
| Mean CDD | 442.13 | 350.74 | 720.63 | 68.41 | 282.05 |
| Stations within 50 miles | 27 | 33 | 41 | 70 | 35 |
| $n$ | 7,515 | 8,693 | 10,895 | 22,393 | 5,770 |
| $R^2$ | 0.01 | 0.008 | 0.01 | 0.02 | 0.03 |
| Distance | 0.07** | 0.18** | 0.0001 | 0.05** | 0.22** |
| Alt. diff. | −0.01** | −0.03** | 0.02** | −0.003** | 0.002 |
| Lat. diff. | 0.31 | 1.82 | 2.35* | −1.62** | −16.06** |
| Long. diff. | 10.05** | 10.93** | 13.98** | −1.23** | 9.48** |
| Constant | 57.90** | 54.71** | 64.00** | 14.61** | 66.29** |

**Note:** Significance at: *10 and **5 per cent levels, respectively regressions is consistently negative and significant, which follows intuitive sense – we would expect heat risk to decrease as elevation increases. The effect of altitude on rainfall payoffs is unclear, as one might expect – rain likely does not consider the altitude of the land on which it is falling. The coefficient attached to distance for rainfall is, with one exception, significant and positive, meaning that as distance increases the difference in the payoffs does too. Or in other words, as distance increases, the payoffs become less accurate. We might expect a similar result for heat, as stations further apart produce more differentiated results, but it seems that temperatures vary continuously throughout a geographic region and the directionality measures are often of more interest.

These relationships may have interesting implications for future efforts to model spatial variability. In effect, the relationship between rainfall and distance is shown to be strong, which indicates that spatial prediction models could have some success. Heat risk, however, is heavily influenced by altitude and spatial prediction models would do well to account for that effect above and beyond the effect of distance.

These results may seem to be providing little beyond the very obvious – heat risk decreases with altitude because of lower temperatures at higher elevations; likewise, rainfall correlations decrease with distance because of the unpredictable, periodic nature of rainfall. However, there is little evidence for other seemingly obvious implications, like the relationship between latitude and heat risk – we would expect that heat risk would decrease with increased latitudes, but in fact only one of the six coefficients is negative and significant. In fact, it is somewhat remarkable how little we can say about the relationship between simple geographic variables and differences in downside risk. It has been assumed by many researchers that it would be possible to provide a statistical solution to the problem of geographic basis risk; these results belie the fact that weather risk may indeed defeat the ability of statistical methods to predict.

**Improving the fit**

There are a few transformations that we can do to improve the fit, which is not a purely academic exercise if our goal is to make out-of-sample predictions for unknown locations. The easiest way to improve the fit of the regression is to include dummy variables for the weather stations and years.

The justification for including dummy variables is thus: it is easy to postulate that each station is to some degree idiosyncratic; these dummy variables are intended to catch the effects of nearby lakes or valleys, or anything else that cannot be captured by the simple geographic variables that we use. The dummy variables for each year isolate the amount of variability in any given year because the dependent variable is strictly positive. This will account for any years in which payout differences were more pronounced. Both of these dummy variable types may also be included in a pricing algorithm as well, although if we are pricing a premium for an unknown location for which there have never been weather observations, we cannot use the variables which account for station idiosyncrasies.

In addition, the regression results presented above indicate that geographic variables do not explain the difference in payouts very well, but there is some evidence that the problem is one of scale. More specifically, this difference equation has no way of distinguishing between payouts which are $0/$200 and $1,000/$1,200. In both cases the dependent variable will be $200, even though they are quite different on a percentage basis.

There are several potential ways to modify the equation to account for this. One method is to move the $P_2$ variable to the right side of the equation, where it may be fit with a regression coefficient. This approach improves the fit markedly but necessitates difficult interpretations of the equation. First, if the coefficient attached to the $P_2$ variable is significantly different than one, it is difficult to interpret what that means, because $P_1$ and $P_2$ are identical in nature and the matter of which one is written first depends only on the ($P_1 \geq P_2$) condition. Second, if we are trying to make an out-of-sample prediction, we cannot assume that the $P_1$ variable will be larger than $P_2$, which may bias the results.

The results of these transformations for all stations and years of data are presented in Table VI. The $R^2$ calculations are highlighted and improve considerably, but it is wise to remember standard caveats on the effects of $R^2$ values when adding dozens, if not hundreds, of variables to the equation. Of course, the greatest effect on the $R^2$ value comes from the addition of a single variable (and the manipulation of the dependent variable) that accompanies moving $P_2$ to the right side of the equation.

Remarkably, the $P_2$ variable is highly statistically significant in all places that it is introduced. The two variables that we identified as causal in the regression equation maintain their significance and sign through all modifications even as all other variables experience widely ranging results.

## Out-of-sample predictions

Table VII shows the results of out-of-sample predictions of payoffs in Ithaca, NY for heat and rainfall using several different types of effects for illustration – first with the simple geographic variables, then including the station and year dummy variables, and finally when moving $P_2$ to the right side of the equation. Also of note is that this

| | Original | Incl. station dummies | Incl. year dummies | Incl. station and year dummies | $P_1$ as Y | All effects |
|---|---|---|---|---|---|---|
| *Heat* | | | | | | |
| DF | 5,948 | 5,889 | 5,875 | 5,816 | 5,947 | 5,815 |
| $R^2$ | 0.05 | 0.23 | 0.54 | 0.61 | 0.68 | 0.87 |
| Distance | 0.04 | 0.06** | 0.01 | 0.09** | 0.04 | 0.07** |
| Alt. diff. | −0.01** | −0.03** | −0.01** | 0.02** | −0.01** | 0.04** |
| Lat. diff. | −15.93** | −7.07 | −11.57** | 35.80** | −17.15** | 35.52** |
| Long | 7.85** | −24.86** | 5.44** | 5.65 | 7.85** | 12.45* |
| Constant | 20.72** | 22.05** | 8.27 | −28.22** | 13.03** | −37.62** |
| $P_2$ | – | – | – | – | 1.12** | 0.72** |
| *Rainfall* | | | | | | |
| DF | 34,841 | 34,690 | 34,768 | 34,617 | 34,840 | 34,616 |
| $R^2$ | 0.01 | 0.18 | 0.11 | 0.24 | 0.26 | 0.44 |
| Distance | 0.13** | 0.10** | 0.12** | 0.10** | 0.13** | 0.09** |
| Alt. diff. | 0.01** | −0.03** | 0.01** | −0.02** | 0.01** | −0.03** |
| Lat. diff. | 3.09** | 8.16 | 3.05** | 10.34 | 3.00** | 17.12* |
| Long | −0.74* | 12.82* | −0.53 | 19.81** | −0.61 | 17.93** |
| Constant | 32.35** | 23.67** | 25.33** | 10.08 | 30.45** | 10.71 |
| $P_2$ | – | – | – | – | 1.41** | 0.42** |

**Note:** Significance at: *10 and **5 per cent levels, respectively

**Table VI.**
Results of transformations in the regression equation

| | Prediction ($) | Difference (%) | Correlation |
|---|---|---|---|
| *Heat* | | | |
| Geo. variables only | 25.27 | − 154.8 | − 0.44 |
| With station and year effects | 32.20 | − 100.0 | 0.46 |
| And moving $P_2$ to right side | 67.30 | 4.3 | 0.90 |
| *Rainfall* | | | |
| Geo. variables only | 37.34 | − 71.8 | 0.43 |
| With station and year effects | 51.33 | − 25.0 | 0.52 |
| And moving $P_2$ to right side | 66.81 | 4.0 | 0.62 |

**Table VII.**
Out-of-sample predictions



prediction was only performed when Ithaca was the station listed first (i.e. the $P_1$ variable), the consequence of which is that the payouts are significantly higher ($64.39 and $64.17) than the long-term averages as presented in Table I ($20.76 and $20.87). Whether this has implications for the end results is an important consideration.

What this shows is that the predictions with geographic variables are not very accurate, but improve with the addition of the station and year effects. The strongest effect is obtained by moving $P_2$ to the right side of the equations, which may make sense – the weather observations are the strongest piece of information we have about prevailing conditions in any given year and by taking the difference we often censor that important piece of information. In any case, it must be said that the geographic variables seem to be useful only in the optimization of an already robust distribution – in any successful prediction presented herein, the "heavy lifting" is done by the station, year, and $P_2$ effects. And in the case of rainfall, this entire exercise has resulted in payouts that are in fact slightly worse than the very simplistic approach taken in Table I of simply averaging payouts for each station within 67 miles of Ithaca.

## The next step

While it may be difficult to propose an index insurance contract based on an average of every weather station within a certain area, even if it is most accurate, the concept of doing so reveals a larger principle. Woodard and Garcia (2008a) write that "portfolio" of derivatives from established derivatives markets in large cities could be a solution to the problem of spatial basis risk. While the authors of this paper would not advocate for using information from stations that are very far away, it may be possible to extend this concept by offering an index insurance contract which is a configured portfolio of local stations. This is because our results strongly indicate that the strongest predictor of a payout at any given station is in fact whether or not there is a payout in a nearby station. Based on the discussion presented in this paper, the portfolio would use weather stations as close as possible and explicitly include elevation and distance into the portfolio selection criteria, if not the other variables which may or may not be significant in any geographic location.

The next step for this research would be to construct a function which could serve to provide guidelines for a "portfolio" of weather index insurance priced to the nearby stations. The amount of index insurance purchased for the portfolio from nearby stations would depend on the similarity of the pertinent geographic characteristics. For example, in a topologically diverse area such as New York's Finger Lakes region, this could mean a station that is quite removed in distance but similar in altitude.

Particularly in regard to the relationship between heat and elevation, the closest station in a geographically diverse region may not always be the most similar. Indeed, researchers such as Ritter *et al.* (2012) have started to investigate multi-site rainfall models for hedging geographic basis risk.

A portfolio of index insurance chosen according to pertinent spatial variables would offer several advantages. First, since the pricing would be done at a known location according to historical burn rates, the price calculated would be straightforward and done in a manner consistent with previous applications of index insurance. This would likely lead to easier adoption of the technology, as insurance companies would not need to build additional safeguards for risk into the model than those that already exist in established methodology. Second, by pricing only at known locations, the concepts behind this method are more transparent to the layperson than a mathematically rigorous treatment designed to predict risk at any given point. It is too true that as the complexity of mathematical instruments increases, the basis risk inherent in the equation decreases, along with comprehension. The portfolio approach is a simple extension of existing methodology and has the advantage of being easier to comprehend and price.

Lastly, this approach would likely take the form of a set of suggestions based on the best available evidence that would be easily modified to include additional information. In other words, this approach would allow the buyer of the insurance to individually tailor their index insurance portfolio according to their perceived risk. The farmer is likely to have the most detailed local knowledge as to prevailing weather patterns and will even have information as to the effects of other geographic characteristics that do not show up in the regression equation, such as mountains, bodies of water, or even the general pattern by which rain falls on their fields. The ultimate source of information as to the minimization of basis risk is with the farmer and the portfolio approach would allow the flexibility for farmers to hedge their weather risk in the manner they best see fit.

The major disadvantage of this method is that it would benefit from a rich series of historical data in the area surrounding the point of interest. Unfortunately, this may preclude a portfolio method from being used in developing countries, which are areas of the world that could benefit greatly from the adoption of this technology. Because of this, further research needs to be done to test the applicability of this so-called “portfolio” approach to pricing weather risk.

## Conclusion

This paper discusses the problem of spatial basis risk for weather index insurance, a basic and fundamental problem in the widespread adoption of the technology, and shows that there is no easy solution at present. Even when putting aside considerations of the weather/yield relationship and the predictive power of the selected index on crop yields, geographic basis risk will likely remain a persistent problem due to complex interactions of weather, space, and geography. Our search for a general principle for pricing risk at unknown locations has indicated that the relationship of risk to geographic variables is complex and highly dependent on local conditions.

The two enduring relationships presented in this research are the relationship of altitude on heat risk and distance on precipitation risk, which may have interesting implications for future efforts at spatial prediction models. Rainfall is proven to be heavily influenced by distance measures, which holds promise for future efforts to model spatial variability, but future efforts to model heat risk should explicitly account for altitude as the predominant variable of interest.

Our results also strongly show that the best predictor of a payout at any given location is the presence of a payout at another location. This implies that an insurance product could successfully be written using nearby weather stations in a so-called "portfolio method". Buying index insurance based on multiple weather stations would have advantages above and beyond a pricing strategy which offers a point estimate for risk at unknown locations. A portfolio of index insurance would offer transparent pricing, and would be as easy to understand as policies written on a single location. Likewise, writing index insurance contracts for multiple stations is a simple way to allow the consumer to configure the product to address their own perceived risks. Future research should explore the possibilities of hedging weather risk on multiple stations, with specific emphasis on optimizing weights for the stations based on geographic characteristics and climatological concerns.

The concepts produced in this research provide insight into the possibilities and challenges present in pricing spatial basis risk. The next step for spatial basis risk research is to compare the accuracy of pricing multiple contracts for a "portfolio" of risk at a single location. This is especially true given the potential pitfalls of the method as described above, such as a lack of consistent historical data. Future researchers will be able to carry on the material presented in this paper by building upon the relationships that we did find to be evident in the struggle to overcome a persistent problem in the adoption of weather index insurance.

## References

Ghiulnara, A. and Viegas, C. (2010), "Introduction of weather-derivative concepts: perspectives for Portugal", *Journal of Risk Finance*, Vol. 11 No. 1, pp. 9-19.

Giné, X., Townsend, R. and Vickery, J. (2008), "Patterns of rainfall participation in rural India", *The World Bank Economic Review*, Vol. 22, pp. 539-66.

Göncü, A. (2011), "Pricing temperature-based weather derivatives in China", *Journal of Risk Finance*, Vol. 13 No. 1, pp. 32-44.

Heimfarth, L. and Musshoff, O. (2011), "Weather index-based insurances for farmers in the North China Plain: an analysis of risk reduction potential and basis risk", *Agricultural Finance Review*, Vol. 71 No. 2, pp. 218-39.

Miranda, M.J. and Gonzalez-Vega, C. (2011), "Systemic risk, index insurance, and optimal management of agricultural loan portfolios in developing countries", *American Journal of Agricultural Economics*, Vol. 93 No. 2, pp. 399-406.

Morduch, J. (2006), "Micro-insurance: the next revolution?", in Abhijit, B., Roland, B. and Dilip, M. (Eds), *What Have We Learned About Poverty?*, Oxford University Press, Oxford.

Paulson, N.D., Hart, C.E. and Hayes, D.J. (2010), "A spatial approach to addressing weather derivatives", *Agricultural Finance Review*, Vol. 70 No. 1, pp. 79-96.

Richards, T.J., Manfredo, M. and Sanders, D. (2004), "Pricing weather derivatives", *American Journal of Agricultural Economics*, Vol. 86 No. 4, pp. 1005-17.

Ritter, M., Musshoff, O. and Odening, M. (2012), "Minimizing geographical basis risk of weather derivatives using a multi-site rainfall model", paper prepared for the 123rd EAAE Seminar, Dublin, February 23-24.

Schlenker, W. and Roberts, M.J. (2006), "Nonlinear effects of weather on corn yields", *Review of Agricultural Economics*, Vol. 28 No. 3, pp. 391-8.

Skees, J.R. (2008), "Innovations in index insurance for the poor in lower income countries", *Agricultural and Resource Economics Review*, Vol. 37 No. 1, pp. 1-15.

Turvey, C.G. (2001), "Weather insurance and specific event risks in agriculture", *Review of Agricultural Economics*, Vol. 23 No. 2, pp. 333-51.

Turvey, C.G. and Kong, R. (2010), "Weather risk and the viability of weather insurance in China's Gansu, Shaanxi, and Henan provinces", *China Agricultural Economic Review*, Vol. 2 No. 1, pp. 5-24.

Turvey, C.G. and Norton, M. (2008), "An internet-based tool for weather risk management", *Agricultural and Resource Economics Review*, Vol. 37 No. 1, pp. 63-78.

Vedenov, D.V. and Barnett, B.J. (2004), "Efficiency of weather derivatives as primary crop insurance instruments", *Journal of Agricultural and Resource Economics*, Vol. 29 No. 3, pp. 387-403.

Woodard, J.D. and Garcia, P. (2008a), "Basis risk and weather hedging effectiveness", *Agricultural Finance Review*, Vol. 68 No. 1, pp. 99-118.

Woodard, J.D. and Garcia, P. (2008b), "Weather derivatives, spatial aggregation, and systemic risk: implications for reinsurance hedging", *Journal of Agriculture and Resource Economics*, Vol. 33 No. 1, pp. 34-51.

Zeng, L. (2000), "Pricing weather derivatives", *Journal of Risk Finance*, Vol. 1 No. 3, pp. 72-8.

## Further reading

Clarke, D. (2011), "Insurance design for developing countries", doctoral thesis, Balliol College, University of Oxford, Oxford.

Hellmuth, M.E., Osgood, D.E., Hess, U., Moorhead, A. and Bhojwani, H. (Eds) (2009), "Index insurance and climate risk: prospects for development and disaster management", Climate and Society No. 2, International Research Institute for Climate and Society (IRI), Columbia University, New York, NY.

Meze-Hausken, E., Patt, A. and Fritz, S. (2009), "Reducing climate risk for micro-insurance providers in Africa: a case study of Ethiopia", *Global Environmental Change*, Vol. 19 No. 1, pp. 66-73.

## Corresponding author

Michael T. Norton can be contacted at: mnorton@iri.columbia.edu

**This article has been cited by:**


1. MußhoffOliver, Oliver Mußhoff, HirschauerNorbert, Norbert Hirschauer, GrünerSven, Sven Grüner, PielstickerStefan, Stefan Pielsticker. Bounded rationality and the adoption of weather index insurance. *Agricultural Finance Review*, ahead of print. [Abstract] [Full Text] [PDF]
2. ChenWen, Wen Chen, HohlRoman, Roman Hohl, TiongLee Kong, Lee Kong Tiong. 2017. Rainfall index insurance for corn farmers in Shandong based on high-resolution weather and yield data. *Agricultural Finance Review* **77**:2, 337-354. [Abstract] [Full Text] [PDF]
3. CastellaniDavide, Davide Castellani, ViganòLaura, Laura Viganò. 2017. Does willingness-to-pay for weather index-based insurance follow covariant shocks?. *International Journal of Bank Marketing* **35**:3, 516-539. [Abstract] [Full Text] [PDF]
4. Apurba Shee, Calum G. Turvey, Joshua Woodard. 2015. A field study for assessing risk-contingent credit for Kenyan pastoralists and dairy farmers. *Agricultural Finance Review* **75**:3, 330-348. [Abstract] [Full Text] [PDF]
5. Jia Lin, Milton Boyd, Jeffrey Pai, Lysa Porth, Qiao Zhang, Ke Wang. 2015. Factors affecting farmers' willingness to purchase weather index insurance in the Hainan Province of China. *Agricultural Finance Review* **75**:1, 103-113. [Abstract] [Full Text] [PDF]
6. Baojing Sun, Changhao Guo, G. Cornelis van Kooten. 2014. Hedging weather risk for corn production in Northeastern China. *Agricultural Finance Review* **74**:4, 555-572. [Abstract] [Full Text] [PDF]
7. Martin Odening, Zhiwei Shen. 2014. Challenges of insuring weather risk in agriculture. *Agricultural Finance Review* **74**:2, 188-199. [Abstract] [Full Text] [PDF]
8. Niels Pelka, Oliver Musshoff. 2013. Hedging effectiveness of weather derivatives in arable farming – is there a need for mixed indices?. *Agricultural Finance Review* **73**:2, 358-372. [Abstract] [Full Text] [PDF]