



# Improving agricultural microinsurance by applying universal kriging and generalised additive models for interpolation of mean daily temperature

Mitchell Roznik<sup>1</sup> · C. Brock Porth<sup>1</sup> · Lysa Porth<sup>1</sup> · Milton Boyd<sup>1</sup> · Katerina Roznik<sup>1</sup>

Received: 16 June 2018 / Accepted: 4 February 2019 / Published online: 22 March 2019  
© The Author(s) 2019

## Abstract

Agricultural microinsurance has the potential to protect farmers against crop loss caused by extreme adverse weather conditions. Microinsurance policies for small-holder farmers are often designed on the basis of weather indices, whereby weather insurance variables are measured at ground weather stations and then interpolated to the location of the farm. However, a low density of weather stations causes interpolation error, which contributes to basis risk. The objective of this paper is to investigate whether agricultural microinsurance can be improved by reducing interpolation error through advanced interpolation methods, including universal kriging (UK) and generalised additive models (GAM) used with land surface temperature, elevation, and other covariates. Results indicate that for areas with a lower density of weather stations, UK with elevation substantially improves air temperature interpolation accuracy. The approach developed in this paper may help to improve interpolation and could therefore reduce basis risk for agricultural microinsurance in regions with a low density of weather stations, such as in developing countries.

**Keywords** Weather insurance · Microinsurance · Basis risk · Remote sensing · Universal kriging · Generalised additive models

## Introduction

Agricultural risk management and insurance are important to the growth of the agricultural sector worldwide because they help producers to stabilise their income and respond to large and unexpected weather risks. A strong agricultural risk management and insurance system in a global context is important for meeting food demand now and in the future and enhancing world food security (FAO 2009). In developing

---

✉ Mitchell Roznik  
mroznik1@gmail.com

<sup>1</sup> University of Manitoba, Winnipeg, Canada



countries farmers often face poverty and constraints such as a lack of understanding of insurance, unaffordable insurance premiums and inadequate access to credit. Agricultural microinsurance has the potential to reduce poverty and stabilise the income of smallholder farmers. Crop microinsurance tends to use an index-based approach, whereby payouts are based on an easily observable index that is correlated to crop yield outcome. Compared to traditional crop insurance, index-based weather insurance approaches have the potential to address some of the difficulties facing smallholder farmers. For example, index-based weather insurance involves relatively low administration costs, and savings on administration could be passed on to farmers through reduced premiums, resulting in more affordable insurance. However, the problem of basis risk, which refers to the mismatch between the indemnity computed from the index and the actual loss experienced on the farm, may partially explain why index-based insurance has not achieved widespread success.

Most of the index-based crop insurance products around the world, including microinsurance products for smallholder farmers, are based on weather indices. Ground weather station measurements (e.g., temperature and rainfall) are the most common source of information for developing index-based weather insurance (Turvey and McLaurin 2012; Mahul and Skees 2007; Makaudze and Miranda 2010). To a lesser extent, satellite-based vegetation or biophysical parameter indices are used to estimate biomass (Baret et al. 2005; Meroni et al. 2014). The underlying hypothesis is that the weather index is highly correlated with crop yield. In the case of ground weather station data, a low density of ground stations can introduce significant basis risk in situations where a farm is located a considerable distance from the next weather station (Collier et al. 2009; Barnett and Mahul 2007). Some alternative indices that overcome the problem of a low density of weather stations are satellite-based vegetation or biophysical indices that do not rely on weather station infrastructure, but consensus has not been reached regarding the extent to which crop yield can be explained by these methods. Other research has examined the combination of vegetation indices, such as the normalised difference vegetation index (NDVI) with ground weather station variables such as temperature (Turvey and McLaurin 2012). To improve the accuracy of agricultural index-based insurance and index-based microinsurance, it is important to help to overcome the issue of basis risk. In the case of a straightforward index-based weather insurance product based on one weather variable such as temperature, relatively simple interpolation methods are often used in practice. A common example of a simple interpolation method is the nearest neighbour (NN) approach, whereby the weather measurement from the station located at the shortest distance from the farm is used in the insurance index. Since measurements are taken at weather stations and not at the insured farm, there is an interpolation error, which contributes to basis risk.

The objective of this paper is to investigate whether agricultural microinsurance can be improved by reducing interpolation error through the use of advanced interpolation methods that integrate satellite information with observations every 1 km, and ground weather station information with observations that vary in geographic distance. The interpolation methods considered include universal kriging (UK) and generalised additive models (GAM) with ancillary information such as satellite land surface temperature, elevation, and other covariates. Performance is evaluated based on each method's



ability to develop daily fine-grain 1 km resolution temperature layers over Alberta, Canada.

UK improves air temperature interpolation at a farm location by optimally weighting weather station air temperature information from nearby weather stations and incorporating other variables to capture weather trends. GAMs provide a flexible regression framework to model non-linearity in the relationship between air temperature and covariates by allowing covariate domains to be split into subdomains to allow for flexibility. Air temperature interpolation error, measured by root-mean-square error (RMSE) is tested on two weather station density scenarios designed to replicate expected conditions in both developed and developing countries. Case A has a higher density of weather stations and is intended to represent developed countries with a more developed infrastructure. Case B has a randomly sampled lower density of weather stations and is intended to represent developing countries with less developed infrastructure. Case A is empirically modelled using the existing weather station network in Alberta, Canada. Case B is empirically modelled using a randomly generated subset of the Alberta network with stations over twice the average distance apart.

The covariates are tested for both cases to determine whether they improve interpolation evaluated by out-of-sample RMSE. The UK and GAMs performance are also compared and evaluated. Model performance is compared to the simpler methods commonly used in existing index-based weather insurance programmes as a benchmark; these include NN and inverse distance weighting (IDW) methods.

Open source data was collected for ten growing seasons from 2007 to 2016 including elevation, land surface temperatures, soil type, and land cover. Remote sensing data was gathered from the National Aeronautics and Space Administration's (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS) project for land surface temperatures (MOD11A1), transformed wetness index (TWI) (transformed MCD43A4), and land cover (MCD12C1).

The remainder of this paper is organised as follows. First, a background on the experience of smallholder farmers is given with an emphasis on the importance of agriculture and the potential role of agricultural insurance. Next, the benefits and challenges of implementing index-based weather insurance are outlined. Some common weather insurance interpolation methods used in agricultural insurance are discussed, followed by an overview of some methods that could improve interpolation accuracy. The data section introduces each data source used in the analysis. Subsequently, the methods are explained with a focus on the methodology of kriging, and the results are presented. The final section gives a summary and discusses possible implications.

## Background and literature review

There is evidence to suggest that smallholder farming households take significant steps to self-insure their crops and protect themselves from income shocks. These methods of self-insurance can be costly and may not even provide sufficient levels of support during extreme and widespread events such as drought (Dercon et al. 2005; Keyzer et al. 2007; Rosenzweig and Binswanger 1992). Some strategies limit



short-term risk but reduce long-term wealth creation. For example, if farmers plant crops with lower but more reliable yields, or if they limit input expenses to reduce the negative shock of crop failure, they are in fact reducing their income. The economic impacts of uninsured weather events may limit households' ability to accumulate wealth (Jensen 2000). Without investment in human capital and wealth accumulation it is difficult to exit the cycle of poverty. Agricultural insurance and other risk management approaches can contribute to improving agricultural productivity by helping producers to invest in more productive agricultural practices (Porth and Tan 2015).

It is well known that agriculture is a risky industry, and farmers may avoid taking risks in the face of extreme loss (Rosenzweig and Wolpin 1993). When farmers limit risk-taking, they often avoid innovations that have the potential to increase productivity. Farmers often rely on their own savings to cushion potential income shortfalls, and investing in technology often requires them to spend money without being certain of a positive return (Dercon 1996). Further, low-income farmers who have borrowed from others may be discouraged from investing funds in higher-reward activities out of fear of the increased default risk; this is described by Boucher et al. (2008) as risk rationing. Indirectly, lenders often charge higher costs for credit to smallholders, which further discourages technology adoption. Also, some banks do not give credit to farmers until they are sure that the conditions are favourable enough to avoid risk to crops already growing, and this creates an obvious shortcoming that is unsuitable for smallholders who need credit to buy inputs at planting time. Lack of access to credit therefore restricts access to agricultural inputs and technologies, such as improved seeds and fertilisers.

Despite the advantages that insurance offers to smallholders, traditional crop insurance has often been difficult to implement outside developed countries. This may be due to moral hazard, whereby farmers who have insured their crop may intentionally reduce the level of crop management or inputs without concern for reduced yield. The problem of adverse selection is also a concern for individual indemnity-based crop insurance, as it may lead to mispricing and the potential for cross-subsidisation of premiums. Further, the high administration costs associated with crop loss adjustment can be difficult to manage, given that developing countries tend to have a large number of small farms.

### **Indexed-based weather insurance**

An alternative to traditional crop insurance is index-based weather insurance, which avoids the critical issues of moral hazard and adverse selection, because the indemnity depends on the performance of an underlying index that is independent of the policyholder's behaviour and cannot be manipulated by the insured customer. Thus index-based insurance is verifiable and transparent. Index-based weather insurance is the most prevalent type, whereby losses are modelled on an underlying index, such as cumulative rainfall or maximum temperature.

In addition to overcoming the challenges associated with moral hazard and adverse selection, index-based insurance may theoretically be more cost efficient,



because claims are paid directly based on the index, removing the need for loss assessments at individual farms and saving administration costs. All this makes index-based weather insurance especially useful in the context of many developing countries where crop loss adjustment is difficult and costly, since the majority of farmers are smallholders with less than two acres of land. Although index insurance offers advantages compared to traditional indemnity-based insurance, a new problem of basis risk arises. Basis risk refers to the situation where the loss computed by the index differs from the loss experienced on the farm. This can include the situation in which a farmer experiences a loss but does not receive a payment, or alternatively where a farmer does not experience an actual loss but still receives a payment. Therefore, reducing basis risk is likely to be necessary in order to successfully implement index-based weather insurance (Carter et al. 2014; Clarke 2011; Elabed et al. 2013; Jensen et al. 2018; Lin et al. 2015; Skees 2008).

Generally, basis risk can be categorised as spatial, temporal or variable. Variable basis risk refers to the situation of incorrect variables, or insufficient variables being used to model the relationship with yield. Temporal basis risk refers to the mismatch between the assumptions used to weight the index over the growing period compared to the actual crop growth over different stages of the growing season. The focus of this paper, spatial basis in the context of index-based weather insurance, arises from differences between the weather variables measured at the ground weather station and the actual weather observed at the farm (Norton et al. 2012).

Input interpolation error increases with the distance from the ground weather station to the farm, and this error leads to basis risk. Therefore, the lower density of weather stations in the developing world, together with the problem of incomplete data collection, presents a major challenge to the successful implementation of index-based weather insurance and microinsurance (Hazell et al. 2010). Developing new infrastructure to reduce gaps in the observational network primarily in developing countries is a priority of the World Meteorological Organization (WMO 2016). However, the development of new infrastructure to increase the density of weather stations can be a costly and slow process. Even when the infrastructure has been put in place, long time series of consistent measurements are often required to design insurance products and compute actuarially fair premiums. Within the constraints of existing infrastructure, reducing spatial basis risk can be challenging. One approach to reducing spatial basis risk is through various interpolation techniques, which can be used to predict measurements at unknown locations. In addition, with the increasing availability of free data, a second approach would be to use advanced interpolation techniques that integrate satellite imagery data and other possible variables to improve the weather estimate. With medium-resolution satellite imagery data, weather estimates can be obtained every 1000 m, compared to ground weather stations that are often located more than 20 km apart (Wan et al. 2015).

In commercial applications it is common for input variable interpolation to be computed based on relatively simple methods such as measuring the weather variable at the nearest weather station or weighting weather measurements from nearby stations. For example, in the Corn Heat Unit programme in Alberta, Canada, air temperature (the main input) is measured either from the nearest weather station or from nearby weather stations weighted by the producer (Alberta



Financial Services Corporation 2018). These relatively simple methods may have higher interpolation error than more advanced methods, and the additional interpolation error may contribute to spatial basis risk. Using more complex methods and considering additional variables that account for the spatial autocorrelation structure may help improve the accuracy of the index. In developing countries with a lower density of weather stations the interpolation error based on simple approaches is likely to be greater than in countries with a higher density of weather stations. Therefore, more advanced methods, such as UK, may be helpful in reducing basis risk.

## Climatology and meteorology methods

Climatology and meteorology literature have proposed more complex interpolation methods to reduce interpolation error. Currently, approaches such as UK, which is the focus of this paper, have not been applied in the context of developing agricultural microinsurance. In climatology, interpolation methods with coordinate information (latitude and longitude) and elevation have been effective in improving air temperature interpolation accuracy (Hijmans et al. 2005).

In addition to elevation, researchers have also found that land surface temperature values ( $L_{ST}$ ) are closely correlated to ground air temperature measurements from weather stations (Mildrexler et al. 2011; Vancutsem et al. 2010). Remotely sensed  $L_{ST}$  provides daily measurements of each pixel of a 1 km by 1 km global grid if quality issues do not prohibit the creation of the  $L_{ST}$  value. Since  $L_{ST}$  and weather station air temperature are closely correlated and  $L_{ST}$  has a higher resolution compared to the density of ground weather stations, researchers have investigated the use of  $L_{ST}$  for improving air temperature interpolation (Mildrexler et al. 2011; Vancutsem et al. 2010; Kilibarda et al. 2014; Parmentier et al. 2015, 2014; Hengl et al. 2012).

Three common methods for daily air temperature interpolation are kriging, GAMs, and geographically weighted regression (GWR) (Parmentier et al. 2015). For a more thorough review of common interpolation methods used in climatology and meteorology see Hengl (2009). Traditional techniques for daily interpolation of air temperature are often referred to as single timescale methods, which involve using input air temperature data from local weather stations, while multi-timescale methods such as climatology-aided interpolation add temporal climate data to improve prediction accuracy. Both of these techniques can facilitate other covariates such as  $L_{ST}$  values to improve accuracy.

In Parmentier et al. (2015) both single timescale and multi-timescale climatology-aided interpolation techniques were tested using kriging, GAM, and GWR methods, and multiple covariates including elevation ( $Elev$ ),  $L_{ST}$  and land cover ( $L_C$ ) were tested in each model. Employing both—single timescale and multi-timescale methods—they found that adding elevation improved interpolation accuracy and adding  $L_{ST}$  improved interpolation accuracy for the summer months. Additionally, variations between the three methods—kriging, GAM, and GWR—were relatively low, although kriging had the lowest RMSE for the multi-timescale approach and



GAM had the lowest RMSE for the single timescale approach. However, GAM and GWR may be more effective with a lower density of weather stations. The study area was the state of Oregon in the U.S.

Kilibarda et al. (2014) investigated air temperature interpolation globally for 2011 by comparing global RMSE produced by spatio-temporal kriging methods with the inclusion of  $L_{ST}$ , elevation information and a topographic wetness layer (TWI). These results were compared to less advanced methods with and without the additional covariates. Spatio-temporal kriging methods are multi-timescale techniques that incorporate temporal information in a kriging framework. Results of this analysis show that global mean, minimum and maximum air temperatures can be predicted with an accuracy between  $\pm 2$  °C for areas with a high density of weather stations and from 2 to 4 °C for areas with a low density of weather stations.

Currently, the techniques of climatology and meteorology are not commonly used for index-based weather insurance applications; however, they may significantly reduce basis risk through minimising interpolation error.

## Data

### Weather station daily temperature

Daily historical mean air temperature measurements from weather stations ( $T_{\text{mean}}$ ) were collected from the government of Canada's climate and weather database. Measurements were selected for the growing season from 1 April to 1 November for the period 2007–2016 for a total of ten growing seasons. Stations located within Alberta were considered for analysis; air temperature was measured in °C (Government of Canada 2018).  $T_{\text{mean}}$  measurements were calculated as the average of the minimum and maximum temperatures measured within the daytime interval. To maximise the air temperature sample size, weather stations with continuous daily mean observations for each growing season were selected for the analysis (see Table 1).

**Table 1** Annual number of weather stations used for analysis

Year	Weather stations (n)	Year	Weather stations (n)
2007	143	2013	191
2008	152	2014	196
2009	158	2015	196
2010	162	2016	196
2011	169	–	–
2012	185	Mean	177



## Remote sensing land surface temperature images ( $L_{ST}$ )

MODIS, operating on the Terra and Aqua satellites, acquires remote sensing information on 36 spectral bands. This information is used to develop many data products including MOD11 land surface temperature ( $L_{ST}$ ) and emissivity products. The MOD11  $L_{ST}$  data is retrieved at 1 km resolution by a split-window algorithm and at 6 km resolution by the day/night algorithm. Several  $L_{ST}$  products are available in varying resolution and time increments. MOD11A1 daily images were used for analysis (Wan et al. 2015). The MOD11A1  $L_{ST}$  images were converted to °C from Kelvin using the conversion method outlined in the MOD11A1 user manual. Missing values present a challenge for satellite remote sensing data, as cloud cover can obstruct data collection. Inverse distance weighting (IDW) was used to interpolate these missing values. MOD11A1 data was collected for the 2007 to 2016 growing seasons corresponding to the weather station air temperature time series.

## Remote sensing transformed wetness index ( $TWI$ )

The  $TWI$  is a soil moisture index derived from transformations of the MODIS MCD43A4 version 6 Nadir Bidirectional reflectance distribution function/Albedo (BRDF/Albedo) product. The MCD43A4 product is generated daily and available globally. It uses a 16 day retrieval period and a weighting algorithm to generate the daily pixel value at a pixel size of 500 m by 500 m (Schaaf and Wang 2015).

The  $TWI$  as outlined by Gumbricht (2016) uses two linear transformed components derived from MCD43A4 image bands and creates an index similar to a normalised difference vegetation index (NDVI). First, the components are made from the MCD43A4 bands; then they are used to make the  $TWI$  index. The first component (Eq. 1) represents the soil line by arranging dark soil reflectance to light soil reflectance, and the second component (Eq. 2) represents open water. The  $TWI$  retains the reflectance value scale factors of the MCD43A4 product.

$$sl = 0.3148(RL - 563) + 0.3209(NIR - 1008) + 0.3595(BL - 147) + 0.3364(GL - 507) + 0.2498(SWIR_a - 1531) + 0.6573(SWIRL_b - 1836) + 0.2471(SWIRL_c - 1699), \quad (1)$$

$$w = 0.1882(RL - 563) + 0.0384(NIR - 1008) + 0.3595(BL - 147) + 0.3501(GL - 507) - 0.3581(SWIRL_a - 1531) - 0.1731(SWIRL_b - 1836) - 0.6621(SWIRL_c - 1699) \quad (2)$$

where RL = red, NIR = near infrared, BL = blue, GL = green, SWIR = short wave infrared.

Once the soil line and water components have been created they can be input as variables in the  $TWI$  algorithm. The  $TWI$  algorithm is defined by a reference





iso-wetness line, and the algorithm is applied using a scale preserving trigonometric rotation. This is combined with a rescaling factor and a calibration factor.

$$TWI = R * \frac{\sin(\beta + 45)(sl + w + a) + \cos(\beta + 45)(-sl + w + a)}{\sin(\beta + 45)(sl - w - a) + \cos(\beta + 45)(sl + w + a) + C} \quad (3)$$

where the reference iso-wetness line is represented as  $a$ , the slope of the iso-wetness line as  $\beta$ , the rescaling factor as  $R$ , and the calibration factor as  $C$ . The values used for the  $TWI$  algorithm are: iso-wetness line  $a$  (2080), iso-wetness line slope  $\beta$  ( $-57$ ), calibration factor  $C$  (7000), and rescaling factor  $R$  (5942). For the purpose of this analysis, the reference values used in Gumbricht (2016) were used as the default values. The calibration factor  $C$ , reference iso-wetness line  $a$  and its slope  $\beta$  can be sampled using reference sites.

### Remote sensing land classification ( $L_C$ )

The MODIS MCD12C1 land cover ( $L_C$ ) product <sup>1</sup> provides land classification maps of three classification schemes: the International Geosphere-Biosphere Programme (IGBP), University of Maryland (UMD), and the Leaf Area Index (LAI). In this analysis we use the IGBP classification scheme, which classifies land cover into 17 categories including cropland, and is of particular interest for agriculture (Friedl and Sulla-Menashe 2015). We consider the effect of cropland on mean air temperatures ( $T_{\text{mean}}$ ) and discard other land covers because they were not well represented in the sample.

### Digital elevation model ( $Elev$ )

The Canadian Digital Elevation Model (CDEM) provides an array of vertical elevation values in a common measurement, metres, and vertical accuracy over the Alberta region of approximately 0–5 m (Natural Resources Canada 2016).

### EcoDistrict soil type ( $S_T$ )

Soil type ( $S_T$ ) is grouped spatially within a national ecological framework. The framework's most specific identifier, the EcoDistrict, is composed of several different attribute classifications including soil type. The attributes are spatially located, and each attribute is derived in unique ways (Agriculture and Agri-Food Canada 2013). The EcoDistrict soil development classification system is based on the Canadian System of Soil Classification. There are a total of 33 classes composed of 31 soil types, a water/rock class, and an urban development class. Soil type was transformed into a binary variable representing Chernozem soil, a fertile agriculture soil, and other. Short descriptions of each covariate are listed in Table 2.

<sup>1</sup> [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mcd12c1\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12c1_v006)



**Table 2** Variable descriptions and identifiers

Symbol	Variable	Explanation
$T_{mean}$	Mean Temperature	Daily mean temperature measured at the weather station
$Elev$	Elevation	Digital elevation model
$L_{ST}$	Land surface temperature	MODIS MOD11A1 daily land surface temperature and emissivity product
$TWI$	Transformed wetness index	Relates transformed MODIS MCY43A4 remote sensing images to soil moisture
$L_C^a$	Land cover	MODIS MCD12C1 IGBP classification scheme
$S_T^b$	Soil type	EcoDistrict derived soil type from Canadian Soil classification scheme
$Lat$	Latitude coordinate	Weather station latitude
$Lon$	Longitude coordinate	Weather station longitude

$T_{mean}$ ,  $Elev$ ,  $L_{ST}$ ,  $TWI$ ,  $Lat$ , and  $Lon$  are continuous variables

<sup>a</sup> $L_C$  is a binary variable, 1 = grassland and cropland, 0 = other

<sup>b</sup> $S_T$  is a binary variable, 1 = Chernozem soils, 0 = other

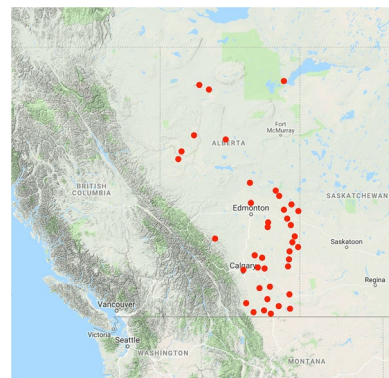
## Methodology

### Weather station density: Cases A and B

Weather station data sets were selected to test interpolation improvements in two cases: Case A (full-density) uses Alberta's original weather station network with an average distance between stations of approximately 20 km; Case B (half-density) uses a random subset of 40 weather stations selected from the same network but with the weather stations over twice the distance apart at an average of 40 km. The lower-density network of Case B was selected to simulate weather station density in developing countries (Figs. 1, 2).



**Fig. 1** Case A: full-density (an average distance between stations of 20 km);  $n = 196$



**Fig. 2** Case B: half-density (with an average distance between stations of 40 km); sample,  $n = 40$



## Data processing and preparation

Before air temperature prediction can be made and accuracy assessed, each variable must be processed and projected on the same coordinate reference system. Analysis was conducted in the NAD83(CSRS)/Alberta 10-TM(Resource) coordinate reference system. Each MODIS data type comes in science data sets that are extracted using the Geospatial Data Abstraction Library (GDAL) tools (GDAL/OGR contributors 2018). One of the conditions required for air temperature interpolation is that all data are projected in the same coordinate space, otherwise the layers would not match interpolation values and an error would result. MODIS data types use a custom sinusoidal grid projection coordinate reference system, and because the weather station data is unprojected, both data types must be projected to match the Alberta grid. A similar transformation must be done for EcoDistrict derived soil type information. These transformations are made using GIS tools available in R statistical software.

Once the target information has been projected on to the specified coordinate reference system, a data set can be extracted from the desired weather station locations for each growing season day for the period 2007–2016. Each data source presents unique challenges, as different formats are used for MOD11A1 land surface temperature ( $L_{ST}$ ), MCD43A4 derived  $TWI$ , MCD12Q1 land cover ( $L_C$ ), soil type ( $S_T$ ), elevation ( $Elev$ ), and weather station mean air temperature ( $T_{mean}$ ). The MOD11A1 derived data sets are represented by gridded layers with land surface temperature values in each corresponding to a 1 km by 1 km grid and  $NA$  values, with pixel values removed due to obstructed measurement or viewing angle error. Similarly, in MCD43A4 the input bands used for the construction of the  $TWI$  values are represented by gridded layers but with a different pixel size of 500 m by 500 m. The land cover data set slightly differs from the other MODIS products; MCD12Q1 comes in a gridded data set with 5600 m by 5600 m pixel size and does not have the same issues with missing values, because land cover is composed annually and the method used to derive land cover uses a larger temporal time window, resulting in less missing values. The EcoDistrict soil type values are stored in spatial polygons with the percentage of soil type coverage and soil type identifiers. In total, 33 soil development types are measured in Canada; however, soil types were aggregated to represent fertile agriculture soil type and other soil types. The soil type for each EcoDistrict was selected as the soil type value. Finally, the digital elevation model presents  $Elev$  values in a gridded layer with an accuracy of approximately 2 to 6 m. These values are transformed on to the Alberta grid and extracted at the weather station locations.

Before the corresponding data values for each data type can be extracted from the weather station locations, the missing values in the MOD11A1  $L_{ST}$  grid layer and the processed MCD43A4  $TWI$  grid layer are filled by IDW. Next, the value for each variable at the intersection of the weather station points is extracted into a spatial point data set. These spatial points are then used for analysis, and the five methods outlined below are used to predict mean air temperature ( $T_{mean}$ ): NN, IDW, OK, UK, and GAMs.



### Nearest neighbour (NN)

The simplest approach to index-based weather insurance is to measure the index variables at the weather stations and base farmers' claims on the nearest weather station to their location. This method is easy to use and transparent for both the insurer and the producer. However, it ignores additional available data, and this may lead to a higher interpolation error than when using methods that incorporate additional information.

### Inverse distance weighting (IDW)

IDW is similar to the NN approach but uses multiple weather stations instead of only one to make a weighted interpolation of air temperature. Weighting is determined based on distance to nearby weather stations and by an a priori decay rate that is defaulted to  $p = 2$ . The predicted air temperature at a given location is then a weighted average of the air temperatures recorded at nearby weather stations.

$$\hat{y} = \sum_{i=1}^n \lambda_i(s_0) \cdot z(s_i), \quad (4)$$

where  $\lambda_i$  is the weight assigned to the air temperature at weather station  $i$ , and the sum of the weather station weights must equal one.

Using the IDW method, weights are determined by distance, and all available weather stations are used as input unless a maximum search radius is pre-determined. A maximum search radius was not set for the analysis.

$$\lambda_i(s_0) = \frac{\frac{1}{d^p(s_0, s_i)}}{\sum_{i=0}^n \frac{1}{d^p(s_0, s_i)}} p = 2, \quad (5)$$

where  $p$  is a coefficient used to adjust weights, and  $d(s_0, s_i)$  is the distance between the target weather station and a known sampled weather station.

### Ordinary kriging (OK)

OK was initially developed for mining applications (Krige 1951; Cressie 2015). It is similar to the previous methods in that it uses weather station location information, but it does not use other covariates in the model and does not include information regarding spatial autocorrelation structure to improve interpolation estimates. Air temperature predictions are based on the following model:

$$Z(s) = \mu + \varepsilon'(s), \quad (6)$$

where  $\mu$  is the constant global mean and  $\varepsilon$  is the spatially correlated stochastic error.

OK predictions are made in the same way as IDW predictions, as shown in Formula (2). Weights are determined to reflect the true spatial autocorrelation structure by analysing the semivariances, where semivariance is defined as:



$$\gamma(h) = \frac{1}{2}E[(z(s_i) - z(s_i + h))^2], \quad (7)$$

where  $z(s_i)$  is air temperature at a select weather station, and  $z(s_i + h)$  is air temperature at a neighbouring weather station at a distance of  $s_i + h$ .

If there are  $n$  unique weather station observations, then  $n*(n - 1)/2$  point pairs are available to compute semivariances. If the semivariances for these point pairs are plotted by distance, we have a variogram cloud that can be binned by distance and further examined. The result of this process is the experimental variogram, which displays the structure of the spatial autocorrelation. Common variogram models include spherical, exponential, and Gaussian models (Cressie 2015; Bivand et al. 2008). Variograms are usually fitted by re-weighted least squares methods, where the weights are determined by the number of point pairs or based on distance (Hengl 2009; Cressie 2015).

Once the variogram is estimated, semivariances are calculated for all the weather station air temperature measurements, and the kriging weights can be solved by:

$$\lambda_0 = C^{-1} \cdot C_0; C(|h| = 0) = C_0 + C_1, \quad (8)$$

where  $C$  is a covariance matrix for  $n \times n$ , and  $C_0$  is a vector of covariances at the new weather station location.

A disadvantage of OK is that it requires all observations to follow an identically and independently distributed stationary mean. This requirement will likely not hold for the Alberta region, because the geographic area is extensive and temperature variation may be large due to factors such as elevation.

### Universal kriging (UK)

UK is an extension of OK that adds explanatory variables to the model to address the stationarity concerns associated with OK. For kriging prediction, variability is separated into global and local components. UK explains the variation in the global component with a trend surface that may include covariates, resulting in residuals that are more likely to follow an individual identically distributed stationary mean. UK requires the fitting of variograms for each day over the time period. The automatic fitting of variograms using the weighted least squares approach was implemented in the automap R package (Hiemstra et al. 2009).

### Generalised additive models (GAM)

GAMs are generalised linear models with a linear predictor composed of a sum of smooth functions of covariates (Hastie and Tibshirani 1990). They are frequently used for interpolation purposes and provide flexibility for applying many different relationships. GAMs require the choice of covariate basis functions and knot parameters. A common basis function choice is regression splines, specifically thin-plate splines, which are often used for interpolation purposes (Parmentier et al. 2014; Hijmans et al. 2005). In this paper they are chosen for basis functions. The choice of knot parameters is determined through a cross-validation approach. For



more information regarding GAMs, refer to Wood (2006). GAMs are fitted using the `mgcv` package available in R using the GCV cross-validation method (Wood 2006).

## Model assessment

Model interpolation performance is assessed using out-of-sample RMSE, mean absolute error (MAE), and mean error (ME) determined by tenfold cross-validation. The model interpolation accuracy is assessed based on RMSE and MAE measured in °C, and the average bias is assessed by ME. As an example, RMSE is outlined below

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{z}(s_i) - z(s_i))^2}{n}}, \quad (9)$$

where  $\hat{z}(s_i)$  is the predicted value at the weather station  $i$ ,  $z(s_i)$  is the observed air temperature value at weather station  $i$ , and cross-validation is performed on the data set for each day over the period 2007–2016.

Interpolation performance is evaluated for Case A (full-density) and Case B (half-density). The analysis follows three steps for each case. First, model performance is evaluated relative to the simpler interpolation methods, including NN and IDW. In this step each covariate is added in a UK and a GAM framework, and model accuracy is compared relative to the model performance over the aggregated time period. Second, model performance is evaluated for the more complicated models that add additional covariates to the models tested in the first step. These models are evaluated based on their performance relative to the best performing model in the first step. Third, the temporal performance of the models is evaluated. The temporal performance of the simpler interpolation methods, which include the NN and IDW approaches, and the top performing advanced interpolation methods, including UK and GAM, are then compared over the aggregated growing season months April to October.

## Results

### Full-density network: Case A

First, the UK and GAM interpolation methods are evaluated relative to the simpler methods, which include the NN and IDW approaches. NN and IDW are commonly used interpolation approaches in existing index-based weather insurance programmes, and more advanced methods may improve interpolation accuracy. In Table 3 model performance is displayed for each model aggregated over each growing season for the period 2007–2016 based on standard assessment metrics that include RMSE, MAE and ME. In Table 3 in mod4 to mod8 for the advanced interpolation methods that include UK and GAM one covariate is added to each to



**Table 3** Results: simple methods and single covariate models

Model name	Model formula	Method	RMSE	MAE	ME
mod1	$T_{\text{mean}} \sim 1^{\text{a}}$	NN <sup>1</sup>	1.0063	0.7431	- 0.06751
mod2	$T_{\text{mean}} \sim 1^{\text{b}}$	IDW <sup>2</sup>	0.9613	0.6997	- 0.08054
mod3	$T_{\text{mean}} \sim Lat, Lon$	OK <sup>3</sup>	0.8092	0.5878	- 0.01764
mod4a <sup>c</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev$	UK <sup>4</sup>	0.6915 <sup>d</sup>	0.5160	- 0.00893
mod5a	$T_{\text{mean}} \sim Lat, Lon + L_{ST}$	UK	0.7792	0.5731	- 0.01677
mod6a	$T_{\text{mean}} \sim Lat, Lon + TWI$	UK	0.7852	0.5751	- 0.01768
mod7a	$T_{\text{mean}} \sim Lat, Lon + L_C$	UK	0.7934	0.5796	- 0.01829
mod8a	$T_{\text{mean}} \sim Lat, Lon + S_T$	UK	0.7902	0.5771	- 0.01859
mod4b <sup>c</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev$	GAM <sup>5</sup>	0.7253	0.5475	- 0.00383
mod5b	$T_{\text{mean}} \sim Lat, Lon + L_{ST}$	GAM	0.8345	0.6201	- 0.00628
mod6b	$T_{\text{mean}} \sim Lat, Lon + TWI$	GAM	0.8418	0.6257	- 0.00691
mod7b	$T_{\text{mean}} \sim Lat, Lon + L_C$	GAM	0.8372	0.6221	- 0.00773
mod8b	$T_{\text{mean}} \sim Lat, Lon + S_T$	GAM	0.8335	0.6204	- 0.00729

Case A,  $n = 177$ . All units are in °C

<sup>1</sup>NN is the nearest neighbour method

<sup>2</sup>IDW is inverse distance weighting

<sup>3</sup>OK is ordinary kriging

<sup>4</sup>UK is universal kriging

<sup>5</sup>GAM is generalised additive models

<sup>a</sup>NN method uses coordinate information represented by 1, and searches for the nearest station

<sup>b</sup>IDW method uses coordinate information represented by 1, and a decaying search window

<sup>c</sup>mod4a uses *Lat*, *Lon* and *Elev* to interpolate  $T_{\text{mean}}$  using the UK method, mod4b uses the same variables and the GAM method

<sup>d</sup>mod4a, UK with *Elev*, is the best performing simple or single covariate interpolation model

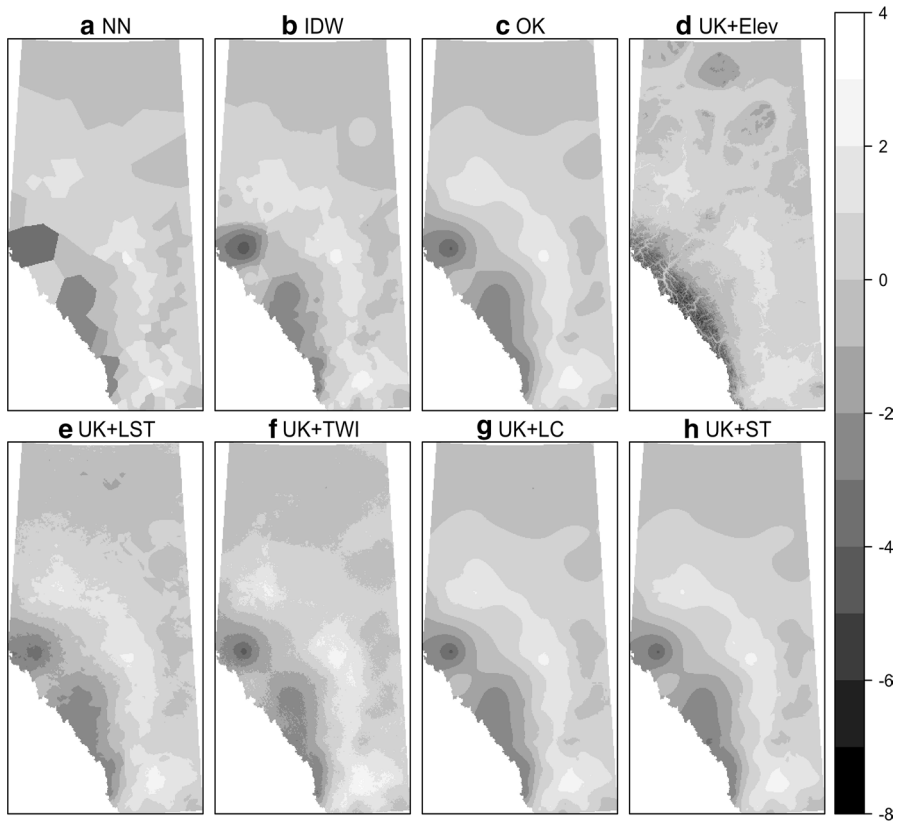
identify the covariate that most improves interpolation accuracy. The covariates tested include elevation (*Elev*), land surface temperature ( $L_{ST}$ ), *TWI*, satellite land cover ( $L_C$ ), and soil type ( $S_T$ ).

The results show that the more advanced methods, including UK and GAM, make substantial improvements in model performance over the simpler interpolation methods of NN and IDW in terms of reducing out-of-sample RMSE. The average RMSE for the NN approach over the time period was 1.0063 compared to the top performing single covariate advanced model UK with *Elev*, which had an RMSE of 0.6915. Further, the UK method seems to outperform the GAM method for each covariate combination.

According to the results, ME seems to be slightly negative for the simpler methods of NN and IDW. GAMs have less negative ME than UK models and far less compared to the simpler methods.

The covariate that seems to have the most impact on interpolation performance over the time period is *Elev*. The addition of *Elev* in both the UK and GAM framework results in the top performing models with RMSE of 0.6915





**Fig. 3** Simple methods and single covariate models: 1 June 2016 surface maps, Alberta, Canada. Case A. **a** NN, **b** IDW, **c** OK, **d** UK and elevation, **e** UK and land surface temperature, **f** UK and transformed wetness index, **g** UK and land cover, **h** UK and soil type. The addition of the elevation covariate in the UK model shown in **d** produces a fine-grain surface that captures the spatial variability. The addition of other covariates does not seem to capture the spatial variability as well as the addition of elevation. Images have been scaled for visual appearance

and 0.7253, respectively. This is consistent with interpolation literature, as the addition of elevation information is believed to improve temperature interpolation accuracy. The effect of adding covariate information for interpolation can be observed visually as in Parmentier et al. (2014), who show that adding important covariates improves the fine-grain characteristics of the predicted surface. A surface map for 1 June 2016 is shown in Fig. 3 as an example to illustrate the fine-grain characteristics of each model as well as the improved texture that each covariate adds. The addition of *Elev* improves the surface visual texture, and additionally the  $L_{ST}$  and  $TWI$  covariates improve the fine-grain surface.

Next, two additional covariate models are assessed to determine whether UK with *Elev*, and GAMs with *Elev* can be improved by adding one of the additional covariates, including  $L_{ST}$ ,  $TWI$ ,  $L_C$ , and  $S_T$ . Table 4 shows the performance results of each of the tested UK and GAM models. The UK models continue to outperform





**Table 4** Results: multiple covariate models

Model name	Model formula	Method	RMSE	MAE	ME
mod1a	$T_{\text{mean}} \sim Lat, Lon + Elev$	UK <sup>1</sup>	0.6915	0.5160	-0.00893
mod2a <sup>a</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev + L_{ST}$	UK	0.6892	0.5173	-0.00693
mod3a <sup>a</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev + TWI$	UK	0.6891 <sup>b</sup>	0.5165	-0.00782
mod4a	$T_{\text{mean}} \sim Lat, Lon + Elev + L_C$	UK	0.6917	0.5177	-0.00918
mod5a <sup>a</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev + S_T$	UK	0.6900	0.5170	-0.00820
mod1b	$T_{\text{mean}} \sim Lat, Lon + Elev$	GAM <sup>2</sup>	0.7253	0.5475	-0.00383
mod2b	$T_{\text{mean}} \sim Lat, Lon + Elev + L_{ST}$	GAM	0.7285	0.5496	-0.00306
mod3b	$T_{\text{mean}} \sim Lat, Lon + Elev + TWI$	GAM	0.7283	0.5493	-0.00304
mod4b	$T_{\text{mean}} \sim Lat, Lon + Elev + L_C$	GAM	0.7263	0.5487	-0.00450
mod5b	$T_{\text{mean}} \sim Lat, Lon + Elev + S_T$	GAM	0.7234	0.5471	-0.00358

Case A,  $n = 177$ . UK outperform GAM models. All units are in °C

<sup>1</sup>UK is universal kriging

<sup>2</sup>GAM is generalised additive models

<sup>a</sup>mod2a, mod3a, and mod5a outperformed mod1a, the UK model with elevation (*Elev*)

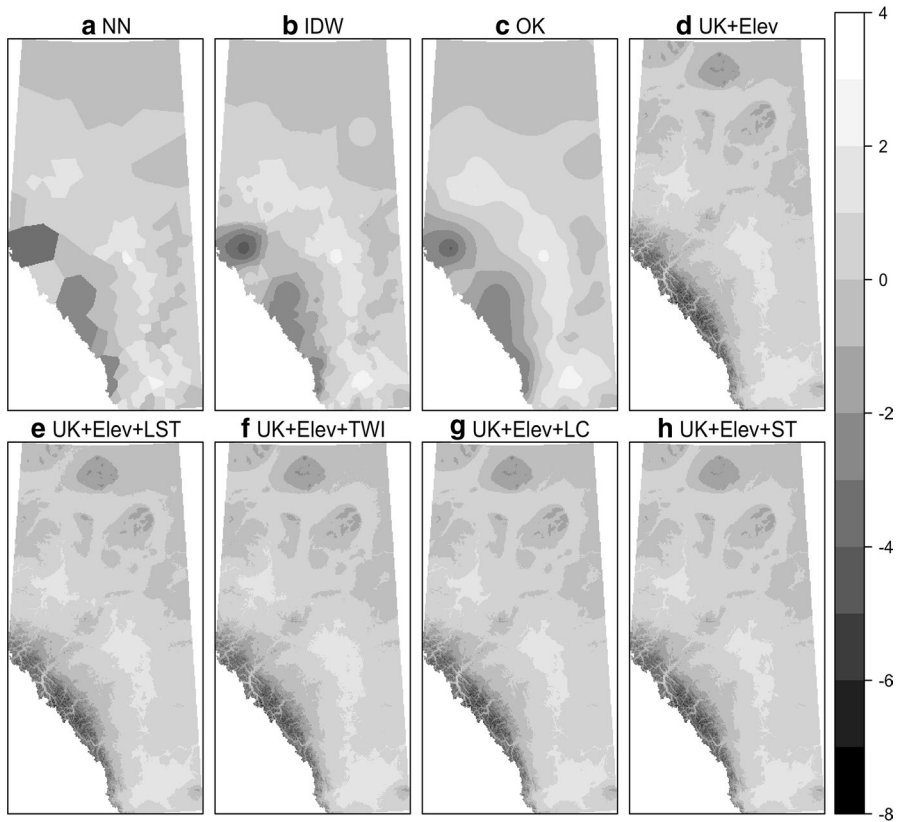
<sup>b</sup>mod3a, UK with *Elev* and *TWI*, was the best performing interpolation model

the GAMs, and the addition of  $L_{ST}$ , *TWI*, and  $S_T$  to the UK model with *Elev* seems to improve interpolation accuracy. The top performing models according to RMSE are UK with *Elev* and *TWI* (mod3), and UK with *Elev* and  $L_{ST}$  (mod2). The improvement of interpolation performance by using the *Elev* and  $L_{ST}$  covariates is consistent with Parmentier et al. (2015), who show that  $L_{ST}$  may improve interpolation accuracy in the summer months. *TWI* also seems to improve interpolation accuracy over the summer months, and similar information may be gained from the *TWI* index and the  $L_{ST}$  index, but further investigation is needed.

Results suggest that the addition of  $L_{ST}$ , *TWI*, and  $S_T$  to the UK model with *Elev* may improve interpolation accuracy over the summer months. Further, examining the results in Table 4, UK and *Elev* (mod1) has an RMSE of 0.6915 and an MAE of 0.5160, and UK with *Elev* and  $L_{ST}$  (mod2) has an RMSE of 0.6892 and an MAE of 0.5173. These results indicate that the addition of  $L_{ST}$  reduces the RMSE but results in an increase in the MAE. A possible explanation is that RMSE penalises large interpolation errors more heavily than MAE, and the addition of  $L_{ST}$  may reduce the large interpolation misses, but it may also slightly reduce the model accuracy. Similar results are found when adding the covariates *TWI* and  $S_T$ ; however, the addition of the satellite  $L_C$  covariate does not improve the interpolation accuracy. Further, according to the ME results the addition of a second covariate reduces the negative bias over the time period.

The effect of additional covariates can be shown visually, as the simpler methods result in a smooth surface that lacks spatial variability, while advanced methods often capture much of the spatial variability resulting in a fine-grain textured surface. The addition of the covariates on the fine-grain spatial variability of the interpolated surface is shown in Fig. 4. It shows that the improvement for capturing





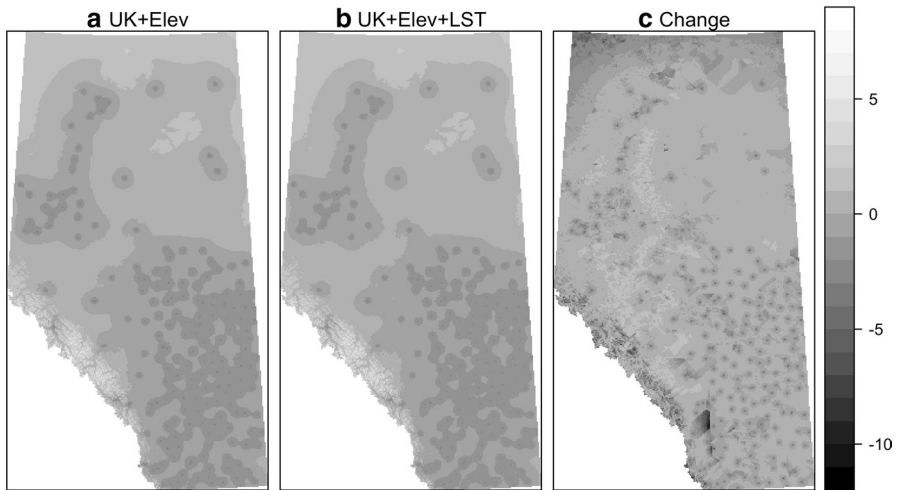
**Fig. 4** Simple methods and multiple covariate models: 1 June 2016 surface maps, Alberta, Canada. Case A. **a** NN, **b** IDW, **c** OK, **d** UK and elevation, **e** UK with elevation and land surface temperature, **f** UK with elevation and *TWI*, **g** UK with elevation and land cover, **h** UK with elevation and soil type. The addition of the elevation covariate in the UK model shown in **d** produces a fine-grain surface that captures the spatial variability. Adding a second covariate to **d** UK with elevation such as land surface temperature, *TWI*, land cover, and soil type does not visually improve the fine-grain texture of the surface. Images have been scaled for visual appearance

spatial variability is large with the addition of *Elev*, but changes are marginal with additional covariates and are not easily recognised through visual assessment.

Further inspection of the change in spatial variability is necessary to assess the impact of adding additional covariates such as  $L_{ST}$ . Figure 5 shows the change in the standard deviation of interpolation accuracy between a. UK model with *Elev*, and b. UK model with *Elev* and  $L_{ST}$ . The positive changes shown in Fig. 5c, displayed in light grey, indicate the regions in which the addition of  $L_{ST}$  reduces the standard deviation of the interpolated surface.

Conversely, the negative changes in Fig. 5c that are displayed in dark grey indicate areas where the addition of  $L_{ST}$  increases the standard deviation of the





**Fig. 5** Change in standard deviation map, 1 June 2016, Alberta, Canada. Case A. **a** UK with elevation, **b** UK with elevation and land surface temperature, **c** change (**a**, **b**). All values are scaled. The standard deviation is mapped for the province of Alberta, Canada for each UK model **a** and **b**. **c** shows the first model **a** minus the second model **b**, and positive values are left where **b** has lower standard deviation. Negative values represent where model **b** has higher standard deviation. Model **b** UK with elevation and land surface temperature has lower standard deviation over most of Alberta, Canada including the major growing areas, and it has higher standard deviation in some of the higher elevation regions where agriculture is less prevalent. Images have been scaled for visual appearance

interpolated surface. According to Fig. 5, the addition of  $L_{ST}$  may improve interpolation accuracy over most of Alberta, including the growing regions, and may reduce the accuracy of interpolation in the mountainous regions where very little agricultural activity occurs.

Temporal interpolation accuracy of each method may change throughout the growing season, and these changes may affect weather insurance performance. Further investigation of the temporal changes of interpolation accuracy over the growing season may be useful for improving index-based weather insurance and micro-insurance programmes. Table 5 displays the monthly out-of-sample interpolation based on RMSE aggregated from 2007 to 2016. Results are for the simpler interpolation methods, which include NN and IDW, and the more advanced methods based on UK, including OK, as well as the four best performing UK models, including UK with *Elev*, UK with *Elev* and  $L_{ST}$ , UK with *Elev* and *TWI*, and UK with *Elev* and  $S_T$ .

According to Table 5, the more advanced interpolation methods Models 3 to Model 7 outperform the simpler methods Models 1 and 2 over each of the aggregated growing season months. Additionally, the UK model with *Elev* and  $L_{ST}$  (Model 5) improves the interpolation performance over UK with *Elev* (Model 4) for the months May, July, August and September. Also, UK with *Elev* and *TWI* (Model 6) improves performance over UK with *Elev* (Model 4) in June, July, August and September. The addition of covariates appears to reduce some of the large interpolation errors. This is best displayed graphically, as shown in Fig. 6.



**Table 5** Temporal interpolation results by RMSE: simple and top performing universal kriging methods

Month	Model 1 <sup>a</sup>	Model 2 <sup>b</sup>	Model 3 <sup>c</sup>	Model 4 <sup>d</sup>	Model 5 <sup>e</sup>	Model 6 <sup>f</sup>	Model 7 <sup>g</sup>
April	1.0403	1.0179	0.7044	0.6990	0.7007	0.7015	0.7001
May	0.9711	0.9887	0.6292	0.6317	0.6305	0.6357	0.6322
June	0.9287	0.9336	0.5993	0.5798	0.5836	0.5939	0.5793
July	0.9441	0.8929	0.6270	0.6260	0.6223	0.6249	0.6294
August	0.9884	0.9227	0.6802	0.6805	0.6768	0.6799	0.6834
September	1.0866	0.9851	0.7872	0.7923	0.7899	0.7905	0.7891
October	1.0836	0.9891	0.8100	0.8114	0.8157	0.8121	0.8122

Case A,  $n = 177$ . The UK models outperform the simpler models NN and IDW over all months, and the addition of a second covariate over Model 4 improves interpolation accuracy in some months. All units are in °C

<sup>a</sup>Nearest neighbour model (NN)

<sup>b</sup>Inverse distance weighting (IDW) model

<sup>c</sup>Ordinary kriging (OK)

<sup>d</sup>Universal kriging (UK) with elevation (*Elev*)

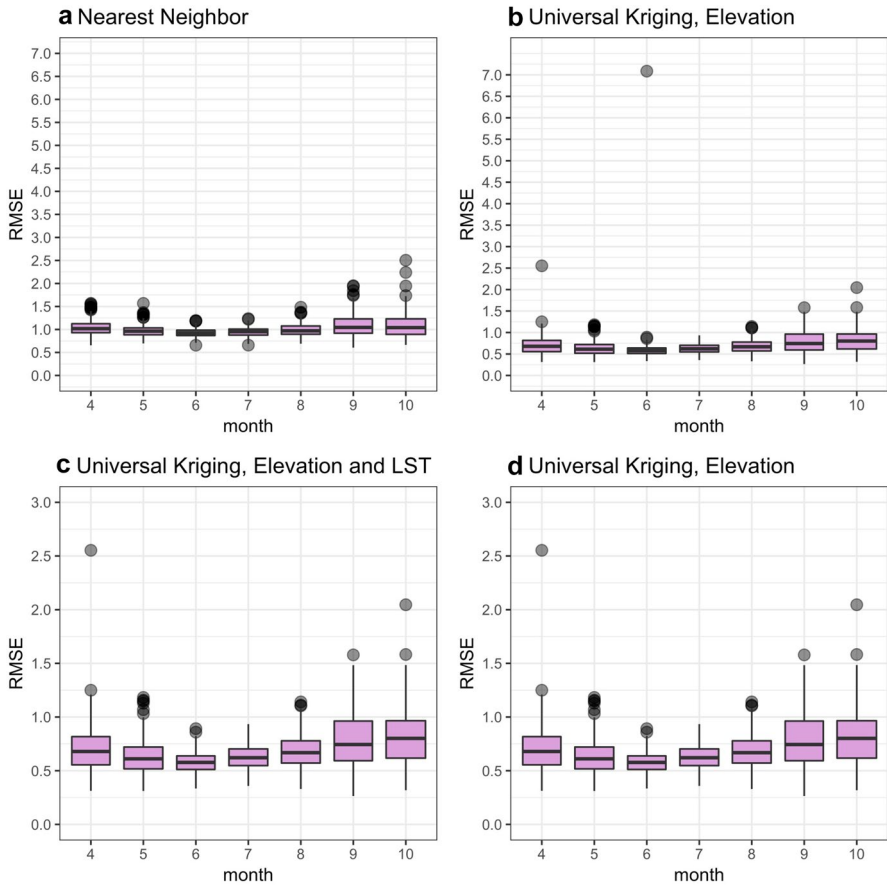
<sup>e</sup>UK with *Elev* and land surface temperature ( $L_{ST}$ )

<sup>f</sup>UK with *Elev* and transformed wetness index (*TWI*)

<sup>g</sup>UK with *Elev* and soil type ( $S_T$ )

Figure 6a shows NN interpolation RMSE throughout the growing season by month for the period 2007–2016. There are concentrations of large interpolation errors, and Fig. 6b shows that using UK with *Elev* significantly improves the interpolation accuracy throughout the season; however, there is a large interpolation error in the month of June. By adding  $L_{ST}$  this large interpolation error is reduced as seen in Fig. 6c. Controlling for large interpolation errors is important for index-based weather insurance purposes. For example, a single large interpolation error could result in a policy failure if a farmer suffered a loss but no loss was signalled by the index.





**Fig. 6** Temporal interpolation results by RMSE for nearest neighbour. Seasonal boxplot. Case A,  $n = 177$ . Comparing **a** and **b**, UK with elevation (*Elev*) significantly improved interpolation accuracy; however, there are some large interpolation misses, specifically in month 6 (June). The addition of land surface temperature ( $L_{ST}$ ) shown in **c** controls for some of the large interpolation misses of **b**. **d** is figure shown with the same scale as **c** for visual accuracy comparison. All units are in  $^{\circ}\text{C}$

### Half-density network: Case B

The half-density network—with twice the average distance between stations as in Case A and intended to simulate developing countries—is examined for each day over the ten growing seasons for the period 2007–2016. The same methods and models are applied to this half-density network as to the full-density network, and model assessment is conducted with the same three-step process described in the section “Model assessment”.

The overall interpolation accuracy of all models and methods decreases as the distance between weather stations increases, as expected. As seen in Table 6, the



**Table 6** Results: simple methods and single covariate models

Model name	Model formula	Method	RMSE	MAE	ME
mod1	$T_{\text{mean}} \sim 1^{\text{a}}$	NN <sup>1</sup>	1.1083	0.8988	-0.05397
mod2	$T_{\text{mean}} \sim 1^{\text{b}}$	IDW <sup>2</sup>	1.0397	0.8385	-0.06546
mod3	$T_{\text{mean}} \sim Lat, Lon$	OK <sup>3</sup>	1.1852	0.9075	0.02049
mod4a <sup>c</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev$	UK <sup>4</sup>	0.8548 <sup>d</sup>	0.6956	0.00415
mod5a	$T_{\text{mean}} \sim Lat, Lon + L_{ST}$	UK	0.9408	0.7677	-0.00850
mod6a	$T_{\text{mean}} \sim Lat, Lon + TWI$	UK	0.9581	0.7792	-0.01571
mod7a	$T_{\text{mean}} \sim Lat, Lon + L_C$	UK	0.9602	0.7782	-0.01153
mod8a	$T_{\text{mean}} \sim Lat, Lon + S_T$	UK	0.9555	0.7757	-0.01419
mod4b <sup>c</sup>	$T_{\text{mean}} \sim Lat, Lon + Elev$	GAM <sup>5</sup>	0.8564	0.6991	-0.00080
mod5b	$T_{\text{mean}} \sim Lat, Lon + L_{ST}$	GAM	0.9495	0.7766	-0.01595
mod6b	$T_{\text{mean}} \sim Lat, Lon + TWI$	GAM	0.9663	0.7899	-0.01721
mod7b	$T_{\text{mean}} \sim Lat, Lon + L_C$	GAM	0.9498	0.7771	-0.01874
mod8b	$T_{\text{mean}} \sim Lat, Lon + S_T$	GAM	0.9412	0.7712	-0.02077

Case B,  $n = 40$ . Despite the lower density of stations in Case B, the top performing model, mod4a, significantly outperforms the simpler models used on the full-density network, Case A. All units are in °C

<sup>1</sup>NN is the nearest neighbour method

<sup>2</sup>IDW is inverse distance weighting

<sup>3</sup>OK is ordinary kriging

<sup>4</sup>UK is universal kriging

<sup>5</sup>GAM is generalised additive models

<sup>a</sup>NN method uses coordinate information represented by 1, and searches for the nearest station

<sup>b</sup>IDW method uses coordinate information represented by 1, and a decaying search window

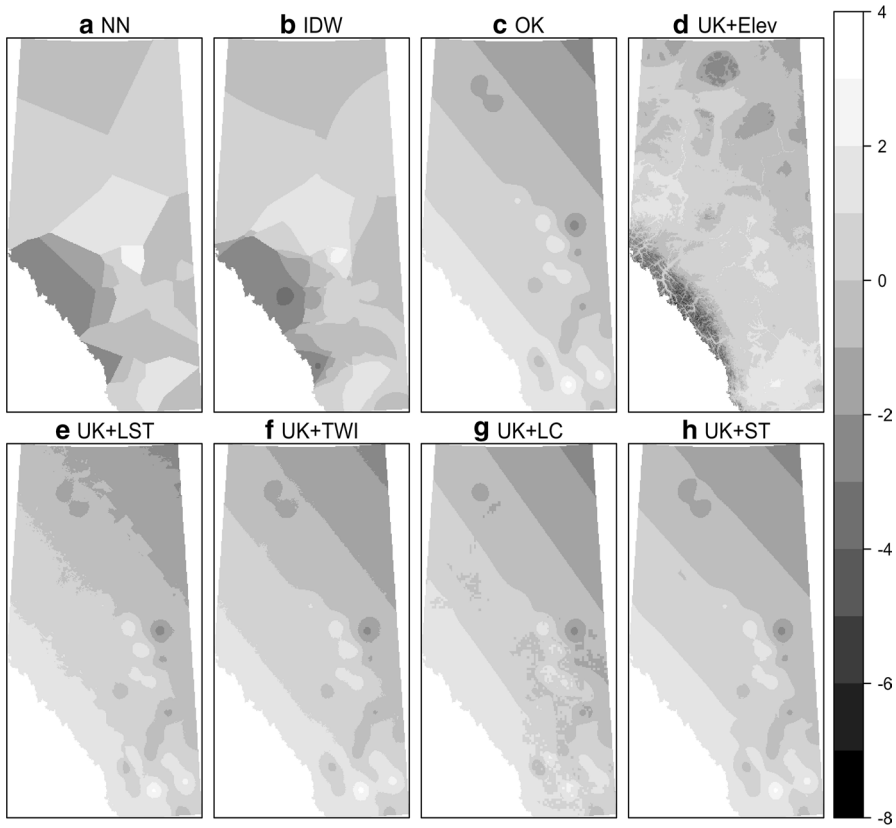
<sup>c</sup>mod4a uses *Lat*, *Lon* and *Elev* to interpolate  $T_{\text{mean}}$  using the UK method, mod4b uses the same variables and the GAM method

<sup>d</sup>mod4a, UK with *Elev*, is the best performing simple or single covariate interpolation model

NN method's RMSE is measured at 1.1083 for the half-density network, compared to 1.0063 for the full-density network. IDW accuracy decreases but not as significantly as in the NN and the OK approaches. OK has an RMSE change from 0.8092 to 1.1852, which is a fairly large decrease in performance; however, with the addition of covariates using the UK model, performance remains high. Consistent with the full-density network, the top performing single covariate model is UK with *Elev* (mod4), which has an RMSE of 0.8548. For comparison, the RMSE of the simpler methods on the full-density network is 1.0063 and 0.9613 for NN and IDW, respectively.

According to the ME results, there seems to be a negative bias for most models with the exception of the OK model (mod3), and the UK with *Elev* model (mod4a). The models with the ME closest to zero are the UK model with  $L_{ST}$  (mod5a), GAM model with *Elev* (mod4b), and the UK model with *Elev* (mod4a).





**Fig. 7** Simple methods and single covariate models: 1 June 2016 surface maps, Alberta, Canada. Case B. **a** NN, **b** IDW **c** OK, **d** UK and elevation, **e** UK and land surface temperature, **f** UK and *TWI* , **g** UK and land cover, **h** UK and soil type. The addition of the elevation covariate in the UK model shown in **d** produces a fine-grain surface that captures the spatial variability. The addition of other covariates does not seem to capture the spatial variability as well as the addition of elevation. Images have been scaled for visual appearance

The effect of adding the covariates can be visually inspected, which shows that adding important covariates may add a texture element to the fine-grain surface, indicating captured spatial variability. As an example, Fig. 7 shows a surface map of Alberta on 1 June 2016 which displays the surface created for the single covariate kriging models, including OK and the simpler methods NN and IDW, using the half-density network.

In Fig. 7 NN and IDW look similar to the surfaces created with the full-density network; however, the OK model Figs. 7c and 7e–h do not seem to capture the spatial variability. In Figure 7d the UK model with *Elev* seems to capture much of the spatial variability structure, resulting in a fine-grain surface.

Next, models with two covariates are assessed to determine if UK with *Elev*, or GAM with *Elev* can be improved by adding one of the additional covariates,  $L_{ST}$ ,



$TWI$ ,  $L_C$ , and  $S_T$ . Table 7 shows the performance results of each tested UK and GAM model. The top performing models assessed by RMSE are UK with *Elev* (mod1a) and UK with *Elev* and  $L_{ST}$  (mod2a). The GAMs do not perform as well as the UK models; however, the model accuracy of the GAMs is closer to the UK methods for the half-density network (Case B) than for the full-density network (Case A).

Results indicate that UK with the *Elev* covariate is the best performing interpolation method, and the addition of a second covariate does not significantly improve interpolation accuracy. Further inspection of method performance results shows that although UK methods outperform GAMs as they did in the full-density case, the difference in performance is much smaller. In Table 7 the RMSE of UK with *Elev* (mod1a) is 0.8548, and GAM with *Elev* (mod1b) is 0.8564. This is compared to the full-density case displayed in Table 6, in which the RMSE of UK with *Elev* (mod1a) is 0.6915, and GAM with *Elev* (mod1b) is 0.7253. ME results suggest that the UK models have a small positive bias and the GAMs have a small negative bias—with the exception of the GAMs with *Elev* and  $L_{ST}$  (mod2b).

**Table 7** Results: multiple covariate models

Model name	Model formula	Method	RMSE	MAE	ME
mod1a	$T_{\text{mean}} \sim Lat, Lon + Elev$	UK <sup>1</sup>	0.8548 <sup>a</sup>	0.6956	0.00415
mod2a	$T_{\text{mean}} \sim Lat, Lon + Elev + L_{ST}$	UK	0.8551 <sup>a</sup>	0.6995	0.00855
mod3a	$T_{\text{mean}} \sim Lat, Lon + Elev + TWI$	UK	0.8602	0.7030	0.00499
mod4a	$T_{\text{mean}} \sim Lat, Lon + Elev + L_C$	UK	0.8644	0.7058	0.00487
mod5a	$T_{\text{mean}} \sim Lat, Lon + Elev + S_T$	UK	0.8603	0.7032	0.00590
mod1b	$T_{\text{mean}} \sim Lat, Lon + Elev$	GAM <sup>2</sup>	0.8564	0.6991	- 0.00080
mod2b	$T_{\text{mean}} \sim Lat, Lon + Elev + L_{ST}$	GAM	0.8759	0.7141	0.00259
mod3b	$T_{\text{mean}} \sim Lat, Lon + Elev + TWI$	GAM	0.8964	0.7288	- 0.00221
mod4b	$T_{\text{mean}} \sim Lat, Lon + Elev + L_C$	GAM	0.8997	0.7294	- 0.00223
mod5b	$T_{\text{mean}} \sim Lat, Lon + Elev + S_T$	GAM	0.8852	0.7210	- 0.00249

Case B,  $n = 40$ . UK outperforms GAM models but with a smaller difference in performance when compared to Case A. All units are in °C

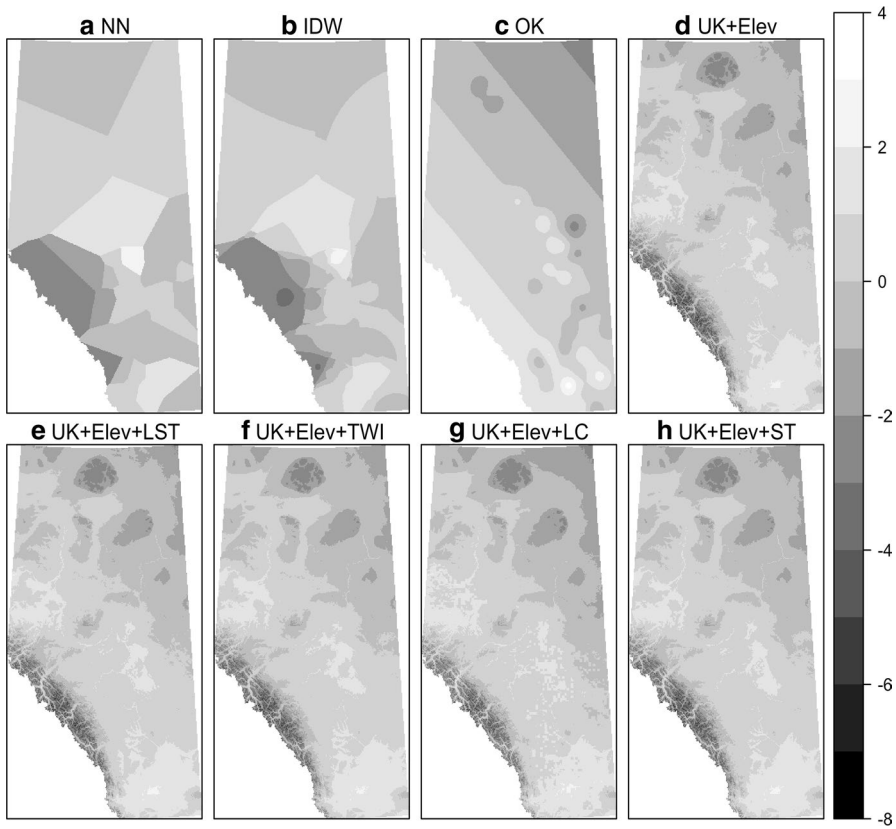
<sup>a</sup>mod1a, UK with *Elev*, is the best performing model; however, the performance of mod2a, UK with *Elev* and land surface temperature ( $L_{ST}$ ), was very similar

<sup>1</sup>UK is universal kriging

<sup>2</sup>GAM is generalised additive models





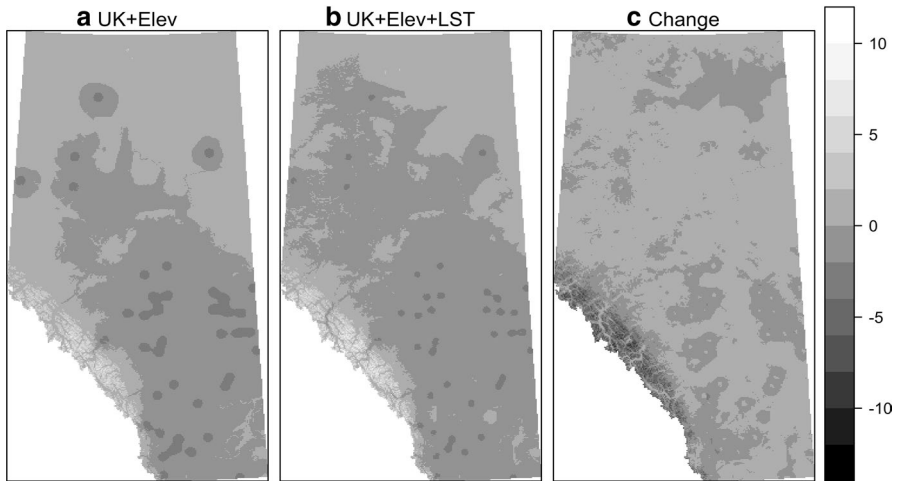


**Fig. 8** Simple methods and multiple covariate models: 1 June 2016 surface maps, Alberta, Canada. Case B. **a** NN, **b** IDW, **c** OK, **d** UK and elevation, **e** UK with elevation and land surface temperature, **f** UK with elevation and *TWI*, **g** UK with elevation and land cover, **h** UK with elevation and soil type. The addition of the elevation covariate in the UK model shown in **d** produces a fine-grain surface that captures the spatial variability. Adding a second covariate to **d** UK with elevation such as land surface temperature, transformed wetness index, land cover, and soil type does not visually improve the fine-grain texture of the surface. Images have been scaled for visual appearance

The effect of adding the covariates can be visually inspected. This shows that adding important covariates may add a texture element to the fine-grain surface, indicating captured spatial variability. Figure 8 shows a surface map of Alberta on 1 June 2016 as an example, displaying the surface created for the kriging models with two covariates, including OK and the simpler methods NN and IDW, using the half-density network.

Through visual inspection of Fig. 8, the addition of *Elev* shown in Fig. 8d improves the fine-grain texture of the surface; however, the impact of the addition of the other covariates on the surface texture displayed in Fig. 8 is not clear.





**Fig. 9** Change in standard deviation map, 1 June 2016, Alberta, Canada. Case B. **a** UK with elevation (*Elev*), **b** UK with *Elev* and land surface temperature ( $L_{ST}$ ), **c** change (**a**, **b**). All values are scaled. The standard deviation is mapped over the province of Alberta, Canada for each UK model **a** and **b**. **c** shows the first model **a** minus the second model **b** and positive values are left where **b** has lower standard deviation. Negative values represent where model **b** has higher standard deviation. Model **b**, UK with *Elev* and  $L_{ST}$  has lower standard deviation over most of Alberta, Canada, including the major growing areas, and it has higher standard deviation in some of the higher elevation regions, where agriculture is less prevalent. Images have been scaled for visual appearance

Further inspection of the change in spatial texture is necessary to assess the impact of additional covariates such as  $L_{ST}$ . Figure 9 shows the change in standard deviation over Alberta for 1 June 2016 as an example, where the positive (light grey) areas show reductions in standard deviation and negative (dark grey) areas show where standard deviation has increased. According to Fig. 9, adding  $L_{ST}$  improves the accuracy of interpolation over most of Alberta, including the major agricultural regions, and reduced interpolation accuracy in the high elevation Rocky Mountains.

The temporal interpolation accuracy of the half-density case B may change throughout the growing season, and these changes may have an impact on the performance of weather insurance. Further investigation of the temporal changes of interpolation accuracy over the growing season may be useful for improving index-based weather insurance and microinsurance programmes. Table 8 displays the monthly out-of-sample interpolation RMSE aggregated for the period 2007–2016 for the simpler interpolation methods including NN and IDW, and the kriging



methods including OK and the four best performing UK models UK with *Elev*, UK with *Elev* and  $L_{ST}$ , UK with *Elev* and *TWI*, and UK with *Elev* and  $S_T$ .

Table 8 shows that the more advanced methods—with the exception of OK (Model 3)—outperform the simpler methods for each month of the growing season. The OK method seems to have difficulty accurately interpolating temperature with the half-density network. Notably for the month of May the OK performance is poor. Further inspection of the interpolation accuracy by month of the more advanced two covariate interpolation models (Models 4 to 7) shows that UK with *Elev* (Model 4) performs best but not in every month. With the addition of  $L_{ST}$  the model UK with *Elev* and  $L_{ST}$  (Model 5) improves interpolation accuracy for the months April, July and August. The addition of *TWI*, UK with *Elev* and *TWI* (Model 6) improves accuracy in the months of July and August. Finally, the addition of  $S_T$ , UK with *Elev* and  $S_T$  (Model 7) improves interpolation accuracy for the month of August. The temporal improvements in some months by adding additional covariates to the UK with *Elev* model are largely due to the reduction of large interpolation errors throughout the growing season. This effect is best represented graphically and can be seen in Fig. 10.

**Table 8** Temporal interpolation results by RMSE: simple and top performing universal kriging methods

Month	Model 1 <sup>a</sup>	Model 2 <sup>b</sup>	Model 3 <sup>c</sup>	Model 4 <sup>d</sup>	Model 5 <sup>e</sup>	Model 6 <sup>f</sup>	Model 7 <sup>g</sup>
April	1.1692	1.1075	1.1034	0.8935	0.8802	0.8937	0.9005
May	1.0967	1.0510	1.7886	0.8003	0.8034	0.8056	0.8041
June	1.0497	1.0118	1.0828	0.7498	0.7612	0.7604	0.7605
July	1.0464	0.9751	1.1209	0.7895	0.7843	0.7837	0.8026
August	1.0515	1.0052	0.9544	0.8486	0.8314	0.8318	0.8392
September	1.1446	1.0421	1.0654	0.9023	0.9208	0.9224	0.9143
October	1.1984	1.0847	1.1727	0.9944	1.0000	1.0183	0.9966

Case B,  $n = 40$ . The UK models outperform the simpler models NN and IDW over all months, and the addition of a second covariate over Model 4 improves interpolation accuracy in some months. Specifically, Model 5, UK with *Elev* and  $L_{ST}$ , outperforms Model 4, UK with *Elev*, in the important growing season months April, July, and August. All units are in °C

<sup>a</sup>Nearest neighbour model (NN)

<sup>b</sup>Inverse distance weighting (IDW) model

<sup>c</sup>Ordinary kriging (OK)

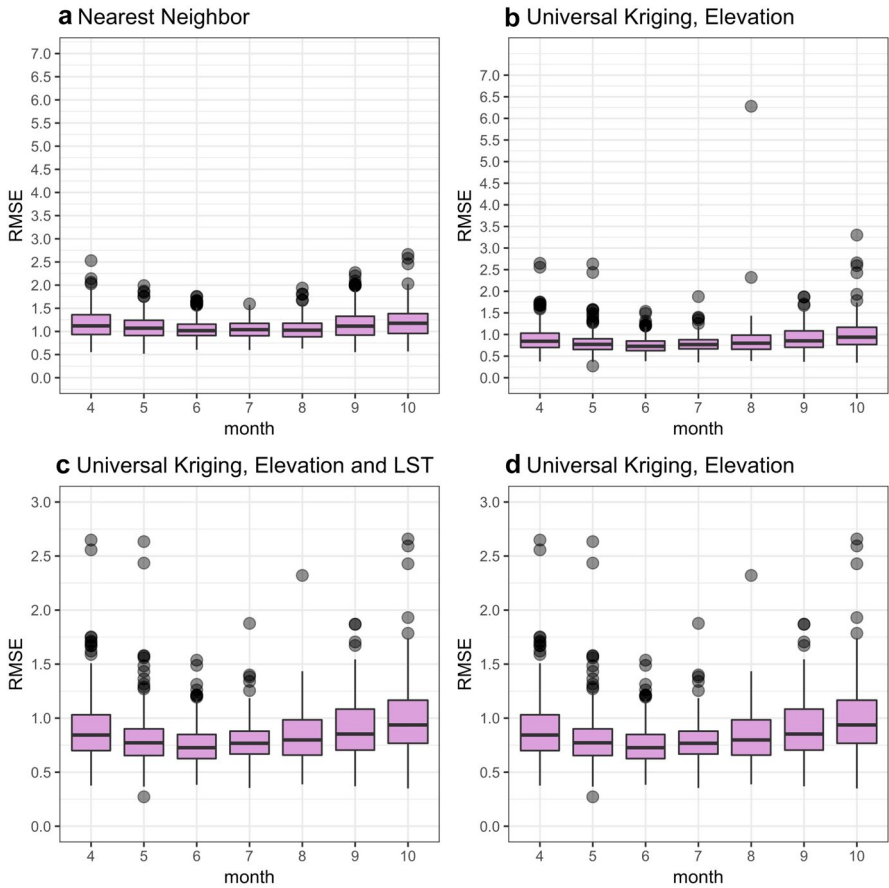
<sup>d</sup>Universal kriging (UK) with elevation (*Elev*)

<sup>e</sup>UK with *Elev* and land surface temperature ( $L_{ST}$ )

<sup>f</sup>UK with *Elev* and transformed wetness index (*TWI*)

<sup>g</sup>UK with *Elev* and soil type ( $S_T$ )





**Fig. 10** Temporal interpolation results by RMSE for nearest neighbour. Seasonal boxplot. Case B,  $n = 40$ . Comparing **a** and **b**, UK with elevation (*Elev*) significantly improves interpolation accuracy; however, there are some large interpolation misses, specifically in month 8 (August). The addition of land surface temperature ( $L_{ST}$ ) shown in **c** controls for some of the large interpolation misses of **b**. Figure 10 is **b** shown with the same scale as **c** for visual accuracy comparison. All units are in  $^{\circ}\text{C}$

Figure 10a shows NN interpolation RMSE throughout the growing season by month for the period 2007–2016. There are concentrations of large interpolation errors, and Fig. 10b shows that using UK with *Elev* significantly improves the interpolation accuracy throughout the season; however, there is a large interpolation error in the month of August. By adding  $L_{ST}$  this large interpolation error can be controlled, as seen in Fig. 10c. Controlling for large interpolation errors is important for index-based weather insurance purposes.



## Summary and implications

Agricultural weather insurance is based on underlying weather variables such as temperature or rainfall. In the case of weather insurance policies for temperature, simple methods are used to interpolate air temperatures observed on the insured farm. Since air temperature measurements are made at weather stations and not at the insured farm itself, there is an interpolation error, which may result in basis risk.

The objective of this paper was to investigate whether agricultural microinsurance could be improved by reducing interpolation error through advanced interpolation methods that integrate satellite information with observations every 1 km, and ground weather station information with observations that vary in geographic distance. The interpolation methods considered include UK and GAM with ancillary information such as satellite land surface temperature, elevation and other covariates. Performance was evaluated based on each method's ability to develop daily fine-grain 1 km resolution temperature layers over Alberta, Canada.

The methods used in this study are UK and GAM interpolation methods. UK improves air temperature interpolation at a farm location by optimally weighting weather station air temperature information from nearby weather stations and incorporating other variables to capture weather trends. GAMs provide a flexible regression framework to model non-linearity in the relationships between air temperature and covariates. Air temperature interpolation error, measured by RMSE, was tested on two different densities of weather stations designed to replicate expected conditions in developed and developing countries. Case A (a full-density network representing developed countries) was empirically modelled using the existing weather station network in Alberta, Canada. Case B (a half-density network representing developing countries) was empirically modelled using a random subset of Alberta's weather station network with the average distance between stations twice that of the full-density network. Data was generated for each day of the analysis.

The covariates were tested for both cases to determine whether they improved interpolation evaluated by out-of-sample RMSE. Also, UK and GAM performances were compared and evaluated. Model performance was compared to the simpler methods commonly used in existing index-based weather insurance programmes that include NN and IDW methods.

Open source data was collected for ten growing seasons from 2007 to 2016, including elevation, land surface temperatures, soil type and land cover. Remote sensing data was gathered from the MODIS project for land surface temperatures (MOD11A1), TWI (transformed MCD43A4), and land cover (MCD12C1).

Results indicate that for both density scenarios the advanced interpolation methods UK and GAM with the addition of covariates substantially improved interpolation accuracy over the simpler methods NN and IDW. For agricultural microinsurance policies in regions where the basis risk caused by interpolation error may otherwise have been too high, these improved interpolation techniques may significantly reduce basis risk.

In Case A (full-density network) UK interpolation with the covariates elevation, and TWI and UK with elevation and land surface temperature resulted in the lowest



interpolation RMSE. By using UK methods with covariate information, air temperature interpolation error may be improved for temperature index-based insurance in cases where weather station density is high. Further analysis suggests that the addition of land surface temperature may help to improve interpolation error over the flat growing regions and reduce interpolation accuracy over the high elevation regions where agriculture is less prevalent. UK outperformed the GAM over the full-density network. These advanced methods may be used in air temperature index-based insurance policies to reduce basis risk caused by interpolation error, such as in the Corn Heat Unit insurance programme in Alberta, Canada.

For Case B (half-density network) UK with the covariates elevation resulted in the top performing model over the ten growing seasons. However, the addition of land surface temperatures led to superior air temperature interpolation accuracy in many months and controlled for large interpolation mistakes. The accuracy achieved by using UK with elevation was similar to a much denser weather station network. Despite the low density of the weather station network in Case B, UK with covariates resulted in higher accuracy than the full-density network in Case A using simpler interpolation methods such as NN and IDW. For temperature weather microinsurance policies in developing countries with low-density networks that result in high interpolation error, the basis risk may be significantly reduced by using UK with elevation information and possibly adding land surface temperature information to reduce the large interpolation errors. Additionally, the use of land surface temperatures may improve air temperature interpolation accuracy over growing regions for important growing season months. In general, the UK method showed better performance than the GAM; however, as distance between weather stations increased, the additional performance improvement gained by using UK over GAM decreased and may continue to decrease as the average distance between weather stations increases. This is consistent with Parmentier et al. (2014), who tested the effect of weather station sparsity on GAM and UK methods, and found that GAMs were less sensitive to weather station sparsity. Although findings suggest that UK methods outperform GAMs for index-based weather insurance programmes in regions where weather station density is low, GAMs with covariates should also be considered and empirically tested.

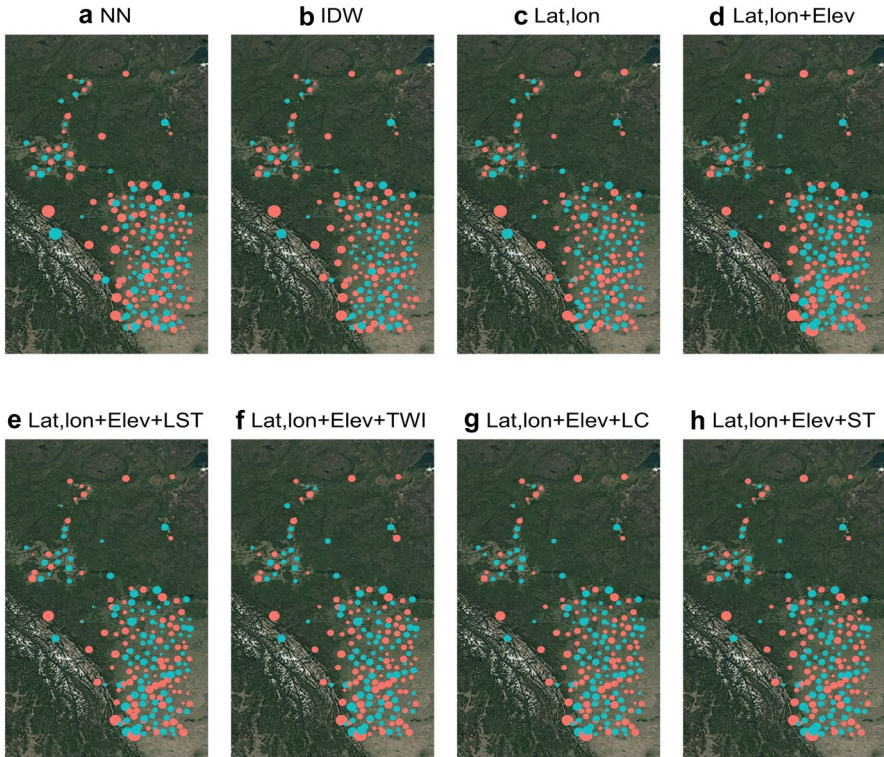
Future research could test whether interpolation error caused by basis risk could be reduced for actual target weather microinsurance areas using the proposed methods. Further, by using actual farm-level data the effect of interpolation error reduction on basis risk could be measured and the effect of basis risk reduction on insurance performance could be simulated. Also, other similar interpolation methods such as GWR or multi-timescale methods such as spatio-temporal kriging could be tested. Additionally, similar studies could investigate the effect of using advanced interpolation methods for reducing basis risk for precipitation-based microinsurance programmes.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



## Appendix

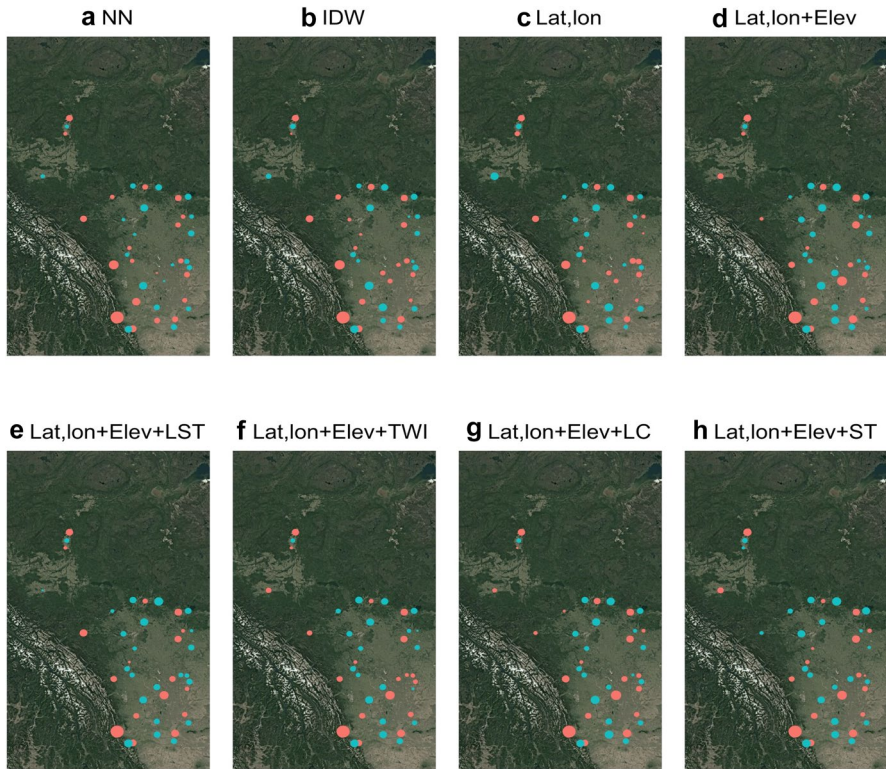
Figures 11 and 12.



**Fig. 11** Model residuals at weather station locations: 1 June 2016 surface maps, Alberta, Canada. Case A,  $n = 177$ . **a** NN, **b** IDW, **c** OK, **d** UK with elevation (*Elev*), **e** UK with *Elev* and land surface temperature ( $L_{ST}$ ), **f** UK with *Elev* and *TWI*, **g** UK with *Elev* and land cover ( $L_C$ ), **h** UK with *Elev* and soil type ( $S_T$ ). Points represent out-of-sample residuals at weather station locations; larger points signify larger residuals. Dark grey points indicate positive residuals and light grey points indicate negative residuals. Visual comparison of model performance is difficult; however, residuals seem to be large in high elevation areas and lower in the growing regions







**Fig. 12** Model residuals at weather station locations: 1 June 2016 surface maps, Alberta, Canada. Case B,  $n = 40$ . **a** NN, **b** IDW, **c** OK, **d** UK with elevation (*Elev*), **e** UK with *Elev* and land surface temperature ( $L_{ST}$ ), **f** UK with *Elev* and *TWI*, **g** UK with *Elev* and land cover ( $L_C$ ), **h** UK with *Elev* and soil type ( $S_T$ ). Points represent out-of-sample residuals at weather station locations, larger points signify larger residuals. Dark grey points indicate positive residuals and light grey points indicate negative residuals. Visual comparison of model performance is difficult; however, residuals seem to be large in high elevation areas and lower in the growing regions

## References

- Agriculture and Agri-Food Canada. 2013. *ISO 19131 National ecological framework for Canada: Data product specification. Revision A*. [http://www.agr.gc.ca/atlas/supportdocument\\_documentdesupport/aafeEcostratification/en/ISO\\_19131\\_National\\_Ecological\\_Framework\\_for\\_Canada\\_Data\\_Product\\_Specification.pdf](http://www.agr.gc.ca/atlas/supportdocument_documentdesupport/aafeEcostratification/en/ISO_19131_National_Ecological_Framework_for_Canada_Data_Product_Specification.pdf). Accessed 10 June 2018.
- Alberta Financial Services Corporation. 2018. *Canada-Alberta AgriInsurance products for 2018 annual crops*. AFSC <https://afsc.ca/wp-nfs/wp-content/uploads/2018/11/Canada-Alberta-AgriInsurance-Products-for-Annual-Crops.pdf>.
- Baret, F., M. Weiss, D. Allard, S. Garrigue, M. Leroy, H. Jeanjean, R. Fernandes, R. Myneni, J. Privette, J. Morisette, et al. 2005. VALERI: A network of sites and a methodology for the validation of medium spatial resolution land satellite products. *Remote Sensing of Environment* 76 (3): 36–39.
- Barnett, B.J., and O. Mahul. 2007. Weather index insurance for agriculture and rural areas in lower-income countries. *American Journal of Agricultural Economics* 89 (5): 1241–1247.





- Bivand, R.S., E. Pebesma, and V. Gomez-Rubio. 2008. *Applied spatial data analysis with R*, vol. 747248717. New York: Springer.
- Boucher, S.R., M.R. Carter, and C. Guirking. 2008. Risk rationing and wealth effects in credit markets: Theory and implications for agricultural development. *American Journal of Agricultural Economics* 90 (2): 409–423.
- Carter, M., A. de Janvry, E. Sadoulet, and A. Sarris. 2014. Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling. *Working Paper P111, FERDI*.
- Clarke, D.J. 2011. *A theory of rational demand for index insurance*. Department of Economics Discussion Paper Series ISSN 1471-0498, University of Oxford.
- Collier, B., J. Skees, and B. Barnett. 2009. Weather index insurance and climate change: Opportunities and challenges in lower income countries. *The Geneva Papers on Risk and Insurance—Issues and Practice* 34 (3): 401–424.
- Cressie, N. 1985. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology* 17 (5): 563–586.
- Cressie, N. 2015. *Statistics for spatial data*. *Wiley series in probability and mathematical statistics*, Rev ed. New York: Wiley.
- Dercon, S. 1996. Risk, crop choice, and savings: Evidence from Tanzania. *Economic Development and Cultural Change* 44 (3): 485–513.
- Dercon, S., J. Hoddinott, and T. Woldehanna. 2005. Shocks and consumption in 15 Ethiopian villages, 1999–2004. *Journal of African Economies* 14 (4): 559–585.
- Elabed, G., M.F. Bellemare, M.R. Carter, and C. Guirking. 2013. Managing basis risk with multiscale index insurance. *Agricultural Economics* 44 (4–5): 419–431.
- Food and Agriculture Organization of the United Nations. 2009. How to feed the world 2050: High-level expert forum. [fao.org/fileadmin/templates/wsfs/docs/Issues\\_papers/HLEF2050\\_Global\\_Agriculture.pdf](http://fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf), accessed 10 February 2018.
- Friedl, M. and D. Sulla-Menashe. 2015. MCD12C1 MODIS Terra and Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006. *NASA EOSDIS Land Processes DAAC*. [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mcd12c1\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12c1_v006).
- GDAL/OGR contributors. 2018. *GDAL/OGR geospatial data abstraction library*. Open Source Geospatial Foundation.
- Government of Canada. 2018. Historical data. [http://climate.weather.gc.ca/historical\\_data/search\\_historic\\_data\\_e.html](http://climate.weather.gc.ca/historical_data/search_historic_data_e.html). Accessed 15 Feb 2018.
- Government of Canada, Natural Resources Canada. 2016. Canadian digital elevation model. <https://open.canada.ca/data/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333>. Accessed 11 Feb 2018.
- Gumbricht, T. 2016. Soil moisture dynamics estimated from MODIS time series images. In *Multitemporal remote sensing*, 233–253. New York: Springer.
- Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized additive models* (1st ed.). *Monographs on statistics and applied probability*. London, New York: Chapman and Hall.
- Hazell, P., J. Anderson, N. Balzer, A. Hastrup Clemmensen, U. Hess, F. Rispoli. 2010. The potential for scale and sustainability in weather index insurance for agriculture and rural livelihoods. Technical report, World Food Programme (WFP).
- Hengl, T. 2009. *A practical guide to geostatistical mapping* (2nd ed.), vol. 52. EUR 22904 EN - 2007. European Commission Joint Research Centre, Institute for Environment and Sustainability.
- Hengl, T., G.B.M. Heuvelink, M.P. Tadić, and E.J. Pebesma. 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology* 107 (1–2): 265–277.
- Hiemstra, P.H., E.J. Pebesma, C.J.W. Twenhöfel, and G.B.M. Heuvelink. 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences* 35 (8): 1711–1721. <https://doi.org/10.1016/j.cageo.2008.10.011>.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25 (15): 1965–1978.
- Jensen, R. 2000. Agricultural volatility and investments in children. *American Economic Review* 90 (2): 399–404.
- Jensen, N.D., A.G. Mude, and C.B. Barrett. 2018. How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya. *Food Policy* 74: 172–198.
- Keyzer, M., V. Molini, and B. van den Boom. 2007. Risk minimizing index functions for price-weather insurance, with application to rural Ghana. *Center for World Food Studies SOW-VU Working Paper*, 07–02.



- Kilibarda, M., T. Hengl, G.B.M. Heuvelink, B. Gräler, E. Pebesma, M.P. Tadić, and B. Bajat. 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres* 119 (5): 2294–2313.
- Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy* 52 (6): 119–139.
- Lin, J., M. Boyd, J. Pai, L. Porth, Q. Zhang, and K. Wang. 2015. Factors affecting farmers' willingness to purchase weather index insurance in the Hainan province of China. *Agricultural Finance Review* 75 (1): 103–113.
- Mahul, O., and J. Skees. 2007. *Managing agricultural risk at the country level: The case of index-based livestock insurance in Mongolia*. Policy Research Working Papers, World Bank Publications.
- Makaudze, E.M., and M.J. Miranda. 2010. Catastrophic drought insurance based on the remotely sensed normalised difference vegetation index for smallholder farmers in Zimbabwe. *Agrekon* 49 (4): 418–432.
- Meroni, M., F. Rembold, M.M. Verstraete, R. Gommès, A. Schucknecht, and G. Beye. 2014. Investigating the relationship between the inter-annual variability of satellite derived vegetation phenology and a proxy of biomass production in the Sahel. *Remote Sensing* 6 (6): 5868–5884.
- Mildrexler, D.J., M. Zhao, and S.W. Running. 2011. A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *Journal of Geophysical Research: Biogeosciences* 116 (G3): 245.
- Norton, M.T., C. Turvey, and D. Osgood. 2012. Quantifying spatial basis risk for weather index insurance. *The Journal of Risk Finance* 14 (1): 20–34.
- Parmentier, B., B. McGill, A.M. Wilson, J. Regetz, W. Jetz, R.P. Guralnick, M.-N. Tuanmu, N. Robinson, and M. Schildhauer. 2014. An assessment of methods and remote-sensing derived covariates for regional predictions of 1 km daily maximum air temperature. *Remote Sensing* 6 (9): 8639–8670.
- Parmentier, B., B.J. McGill, A.M. Wilson, J. Regetz, W. Jetz, R. Guralnick, M.-N. Tuanmu, and M. Schildhauer. 2015. Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon. *International Journal of Climatology* 35 (13): 3862–3878.
- Porth, L., and K.S. Tan. 2015. Agricultural insurance—more room to grow? *The Actuary Magazine* 12 (2): 34.
- Rosenzweig, M.R. and H.P. Binswanger. 1992. *Wealth, weather risk, and the composition and profitability of agricultural investments*, Policy Research Working Paper WPS 1055. World Bank Publications.
- Rosenzweig, M.R., and K.I. Wolpin. 1993. Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India. *Journal of Political Economy* 101 (2): 223–244.
- Schaaf, C. and Z. Wang. 2015. MCD43A4 MODIS Terra and Aqua BRDF/Albedo Nadir BRDF Adjusted RefDaily L3 Global 500m V006. *NASA EOSDIS Land Processes DAAC*.
- Skees, J.R. 2008. Innovations in index insurance for the poor in lower income countries. *Agricultural and Resource Economics Review* 37 (1): 1–15.
- Turvey, C.G., and M.K. McLaurin. 2012. Applicability of the normalized difference vegetation index (NDVI) in index-based crop insurance design. *Weather, Climate, and Society* 4 (4): 271–284.
- Vancutsem, C., P. Ceccato, T. Dinku, and S.J. Connor. 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment* 114 (2): 449–465.
- Wan, Z., S. Hook, and G. Hulley. 2015. MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set]. *NASA EOSDIS LP DAAC*. <https://doi.org/10.5067/modis/mod11a1.006>.
- Wood, S.N. 2006. *Generalized additive models: An introduction with R*, Second ed., Texts in statistical science Boca Raton, FL: Chapman and Hall.
- World Meteorological Organization (WMO). 2016. Global climate observing system: Implementation needs. Technical report. [https://library.wmo.int/index.php?vl=notice\\_display&id=19838#.XGg8TPNKi8o](https://library.wmo.int/index.php?vl=notice_display&id=19838#.XGg8TPNKi8o).



**Mitchell Roznik** is a Ph.D. candidate in Actuarial Science and Agricultural Economics at the University of Manitoba, Canada. His research focus is the application of satellite remote sensing data for agriculture insurance. He holds a bachelor's degree in Finance and a master's degree in Agriculture Economics from the University of Manitoba.

**C. Brock Porth** is an Assistant Professor in Biosystems Engineering, Agricultural and Food Sciences, University of Manitoba. He holds a Ph.D., M.Sc. and B.Sc. in Mechanical Engineering specialising in the areas of Material Science and Aerospace, University of Manitoba. He is a practising Professional Engineer with the provincial licensing Association of Professional Engineers and Geoscientists of Manitoba (APEGM).

**Lysa Porth** is an Associate Professor and Guy Carpenter Professor in Agricultural Risk Management and Insurance and Director of the Warren Centre for Actuarial Studies and Research, University of Manitoba. Lysa also holds a cross appointment in the Department of Agribusiness and Agricultural Economics at the University of Manitoba and is an Adjunct Professor in the Department of Statistics and Actuarial Science at the University of Waterloo. She holds a Ph.D. in Actuarial Science and Agricultural Economics from the University of Manitoba.

**Milton Boyd** is a Professor in the Department of Agribusiness and Agricultural Economics at the University of Manitoba in Winnipeg, Canada. He is also an Adjunct Professor in the area of Risk Management and Actuarial Science at the Asper School of Business of the University of Manitoba. He holds a doctorate degree from Purdue University (USA). Some of his research interests include quantitative finance, risk management, and derivatives.

**Katerina Roznik** is an ecology and mathematics student at the University of Manitoba, Canada, interested in graduate research related to microinsurance, statistics, and environmental sustainability.

