Predicting the yield of spring wheat by weather indices in Finland: implications for designing weather index insurances

Kyösti Pietola[1], Sami Myyrä[1], Lauri Jauhiainen[2] and Pirjo Peltonen-Sainio[2]

[1]*MTT Agrifood Research Finland, Economic Research, Economics and Social Sciences, Latokartanonkaari 9, FI-00790 Helsinki*

[2]*MTT Agrifood Research Finland, Plant Production Research, Jokioinen, FI-31600 Jokioinen*

e-mail: kyosti.pietola@mtt.fi

This paper quantifies the spring wheat yield conditional on temperature- and rainfall-based weather indices in Finland. The estimating equations are standardized and simplified so that they provide information for designing tradable contracts. A simple basket of weather indices, consisting of growing degree days, night frost and rainfall measures, has the potential to hedge about 38% of the wheat grower yield risk, with the remaining 62% being left as uninsured basis risk. Our results have several important implications for the design of simple and tractable weather index-based insurance contracts. The data suggest that the marginal products of weather events have a large variation across time and they are the most significant within certain critical time periods. Therefore, the weather events triggering the indemnity payments should be bounded within certain critical time regimes over their distribution along the growing season.

Key words: Weather index, growing conditions, temperature, precipitation, frost, crop insurance, wheat, yields, instability, risk management

Introduction

Index-based crop insurance contracts have been proposed as the sollution to the informational inefficiencies and problems prevailing in traditional crop insurances. In index-based insurance contracts, the indemnity payments are functions of certain indices, such as an area yield index or weather index (e.g. Van Asseldonk and Oude Lansink 2003). A strength of these contracts is that they do not suffer from adverse selection and moral hazard problems, because the insured agents cannot influence exogenous index values that trigger the indemnity payments

Pietola, K. et al. Predicting yield of spring wheat by weather

(Karuaihe et al. 2008, Barnett et al. 2005, Miranda and Vedenov 2001). Index insurances also maintain the market-driven economic incentives for informal self-insurance mechanisms and investments in risk mitigation measures, such as drainage, irrigation and protection from soil erosion[1]. It is also fair to claim that information about the distribution of future index values is equal amongst the insured farmers and the insurers, and the problems of asymmetric information are not therefore apparent, as in traditional crop insurances. Index-based insurances are likely to have particular strengths in addressing systemic risks, but the existing literature is rather inconsistent about these claims (Xu et al. 2009).

Examples of index-based insurances designed in agriculture include the US programme Group Risk Plan, which is based on an area yield index (Barnett et al. 2005). Weather-based index insurances have also been designed for agricultural purposes in the context of developing countries, where the auditing of household-specific losses incurs high cost (e.g. Barnett and Mahul 2007). Index-based contracts and weather derivatives, in particular, have also received considerable attention in the financial market, and the trade for them has been increasing. Currently, CME Group[2] (2010) offers weather futures and options in four categories: temperature (16 contract types), hurricanes (3 contract types), frost (2 contract types) and snowfall (2 contract types). The temperature-related products are offered for 42 cities throughout the world. Ten of these cities are in Europe, but none in Finland. The cities closest to Finland for which these products are quoted are Stockholm and Oslo. However, neither of these cities can be considered to represent weather in Finland accurately enough, and cannot therefore provide a weather measurement point relevant for hedging against yield risks in Finland.

Although weather index-based crop insurances have many advantages, there is concern about how much the actual yields respond to weather conditions during the growing season or specific weather events, such as the accumulation of temperature, rainfall, frost or hail, and how much is left as unexplained yield variation and to the basis risk (e.g. Peltonen-Sainio et al. 2009a; 2011). Thus, to design attractive and efficient weather-based index insurances we need empirical information on how the yields respond to weather conditions that can be easily and precisely measured and applied in the insurance markets. Information is currently available about linkages between weather and yields, but not in such a simple form that it could be applicable in designing efficient weather indices and insurance contracts in order to protect Finnish grain growers against risks and uncertainty caused by yield variability and loss. Therefore, this paper estimates the response of spring wheat yield to certain weather indices for four different locations in Finland. The estimating equations are standardized and simplified so that they can provide information, such as piece-wise linear marginal products of certain weather conditions, for the design of tradable contracts. Further, the accumulation of weather data is fixed to start on a specific calendar date, and high frequency daily weather data are aggregated across time in a standardized way to guarantee that the results can be used for designing tradable weather indices.

Spring wheat is used as the reference crop, because its role in Finnish cropping systems has markedly increased in recent decades: the number of hectares under wheat has doubled since 1995, with an average annual increase of 6.7%. Furthermore, wheat has generally higher risks for yield losses due to being the latest maturing cereal grown in Finland, i.e. it is harvested later than spring-sown barley and oats, and winter rye (Peltonen-Sainio et al. 2009b). However, this does not necessarily indicate that Finnish farmers have been willing to take more risks since EU accession in 1995. There is in fact some justification for an increase in risk aversion behaviour (Kondouri et al. 2009). One reason for increasing the area under wheat cultivation relates to price risk aversion behaviour, since wheat has remained in EU price intervention schemes, whereas barley and oats have not.

1 The market refers here to the commodity market that is also incomplete, as there is no market for risk and insurance is not available.
2 The Chicago Board of Trade (CBOT) and NYMEX have merged to form CME Group.

We measure both the yield and weather parameters in the same location, and the results indicate the potential of using weather indices for predicting yields on the local scale. Our results provide information that is needed as a first step in designing weather index-based insurance contracts to protect against yield losses. The estimates also shed a base for future research in estimating the relationship between the yields and the weather when the spatial distance between the production site and the weather station increases.[3]

Data

Weather data

The weather data are from the Finnish Meteorological Institute (FMI) over the period from 1970 to 2008. These data consist of the daily accumulation of "growing degree days" (GDD)[4], the daily minimum temperature and precipitation measurements.

The daily data are first aggregated into weekly intervals, and we later test how these weekly data can be further aggregated for being applicable to build index based insurance contracts. For temperature measures (GDD and minimum temperature), the three last weeks are omitted in the analyses, because the accumulation of GDD after the 17th sample week (after August 27) no longer contributes to the yields. For rainfall the situation is different, since excessive rainfall at harvest can still result in severe crop damage (Peltonen-Sainio et al. 2009b).

Growing degree days (GDD)

The growing season comprises the period when the average daily temperature remains above +5 °C. The annual average GDD for our field experimental

data decreases from the south of Finland (1,123 ºC) to the north (1,028 ºC), whereas the standardized maturation GDD's for the wheat cropping season are between 941 and 1,056 ºC, depending on the breeding line. The growing period from sowing to maturation required for spring wheat varieties grown in Finland varies within the range of 100.9 to 109.7 days.

At FMI statistics, the accumulation of growing degree days starts from the beginning of the growing season, which usually takes place in late April in Southern Finland, early May in Central Finland, and the latter half of May in Lapland. However, in this study we have defined the growing season to start each year on 1 May and to continue for 20 weeks (140 days) until 17 September.[5] This approach of locking the growing season to exact calendar dates is often used in corresponding applications, and is referred to as the "biofix" (Xu et al. 2009).

Thus, our weather data over 140 days each year cover the growing season of wheat, even if the true sowing time has annual variation. It also covers the conditions at harvest when risks for yield losses are typically high (Peltonen-Sainio et al. 2009b). Within the typical wheat production areas, located from southern to central coastal regions of the country, the length of the growing season and the accumulation of GDD exceed, on average the requirements for growing wheat.

An alternative approach would have been to start the growing season each year at different time, such as when the accumulated GDD and other measures pass certain threshold. However, even though it is recorded in the Finnish agricultural statistics, the true sowing date is endogenous and could not be used as an exact predetermined base for tradable weather index insurances and derivatives. The tradability of contracts requires that the underlying weather measures that trigger the indemnity payments are precise and locked into exact calendar days.

The GDD index is calculated as the cumulative sum of those daily average air temperatures that exceed five degree Celcius. The daily aver-

---

3 The spatial correlations of weather indices, yields and indemnity payments are estimated in Myyrä et. al (2011).
4 Growing degree days (GDD) are also referred to as the effective temperature sum (Mukula and Rantanen 1987).

---

5 The data are available from the authors upon request.

age is computed using eight measurements sequenced in three hour intervals of equal duration within each day. Formally, the GDD measure is (Kangas et al. 2009):

$$GDD_{\ell,t} = \sum_{t=1}^{T} \max\left(\frac{(^{\circ}C_{\ell,t}^{00} + {}^{\circ}C_{\ell,t}^{03}... + {}^{\circ}C_{\ell,t}^{21})}{8} - 5 , 0 \right)$$

(1)

where subscript ℓ refers to location and

$${}^{\circ}C_{\ell,t}^{00}, {}^{\circ}C_{\ell,t}^{03}, {}^{\circ}C_{\ell,t}^{06},..., {}^{\circ}C_{\ell,t}^{21}$$

refer to the air temperature measurements in degree Celcius scale at three hour intervals, starting at midnight ($^{\circ}C_{\ell,t}^{00}$) and ending at nine o'clock in the evening ($^{\circ}C_{\ell,t}^{21}$). The first day ($t=1$) is fixed at May 1 and the   last day at September 17. Thus the total number of days (T) over which the GDD is computed is 140.

Truncating our GDD accumulation at May 1st did not result in much losses of generality in the current data since the GDD did not accumulate to a large extent before May (Figure 1). The weekly GDD measures gradually increase during the first eleven weeks and then, for the last seven weeks, turn to a decreasing trend. The variation of the weekly GDD is larger up until midsummer (May-June) than in late summer (July-August). On average, May accounts to only 7 – 11% of the total GDD accumulated over the 20-week period.

Daily precipitation

Daily precipitation (RAIN) is measured in millimetres (Figure 2). This index addresses both the drought risk and excessive rain. When the RAIN index is aggregated into weekly intervals, it has either positive or negative impacts on the yield, depending on the time period.

The distribution of precipitation across the growing season is not optimal for crop growth (Peltonen-Sainio et al. 2009b). The rainfall is too scarce in early summer and it increases gradually towards July, also remaining high at harvest. The variability in rainfall also increases from May until July and remains high thereafter. The highest rainfall events are typically observed in July and early August.

The aggregation of data into weekly intervals is reasonable, because rain, even when heavy on a Finnish scale, does not usually destroy the whole harvest at once unless in the form of hail. Our data did not include any two-day rains exceeding 100 mm, and the frequency for rainfall of more than

Fig. 1. The distribution of weekly (week 1 starts at May 1) GDD's during 1970–2008. The six regions (ℓ) are pooled. Tukey box: 50% of observations are within the box, with the lower and upper boundaries of the box being quartiles 1 and 3, respectively. The box is divided by the mode (Q2), and "+" indicates the mean. Circles indicate unusual observations.
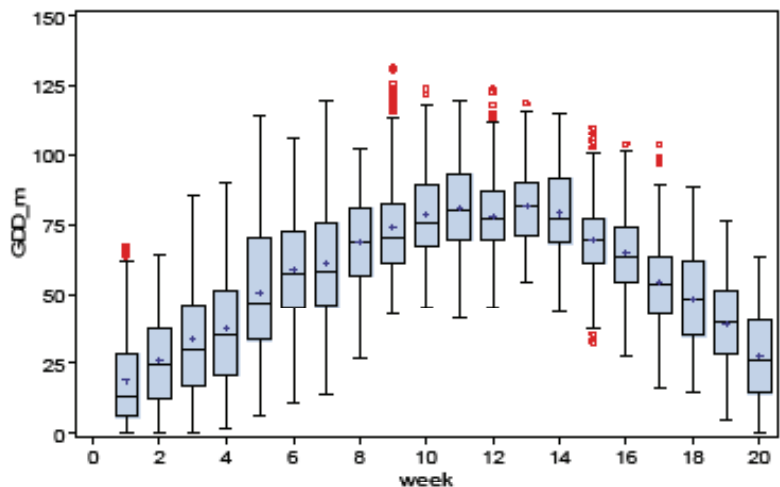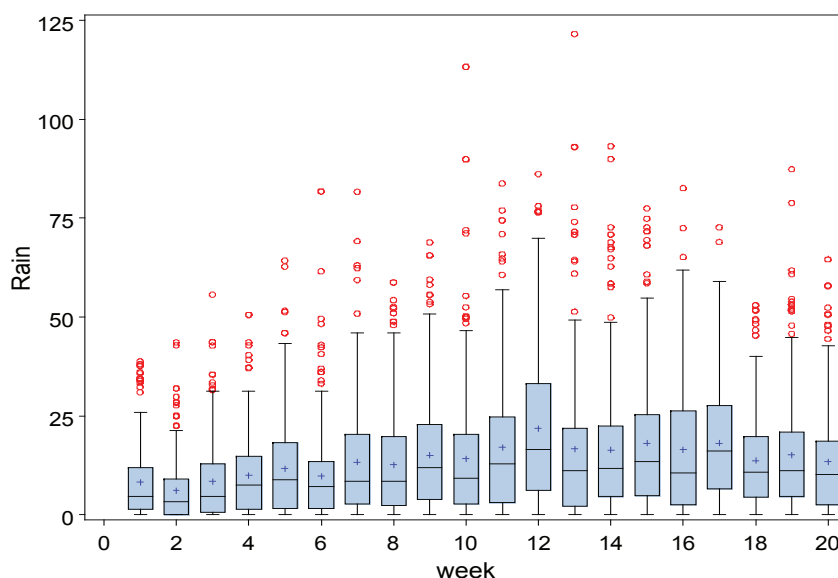
Fig. 2. The distribution of weekly rainfall (in millimetres) over the growing season in 1970-2008. The six sample regions are pooled. Tukey box: see Fig. 1 for details.

80 mm was also too small to identify heavy rains within a shorter than weekly interval . Heavy rains cause lodging. If rains are long-lasting and abundant soils become wet and do not carry heavy machinery. Also harvesting is hampered and cereal quality deteriorated, especially due to the proliferation of microflora that are favoured by wet conditions (Peltonen-Sainio et al. 2009b). The weekly aggregates are also useful in the determination of drought effects as, for instance, rather short periods of drought prior to heading reduce the number of set grains and cause yield penalties that largely cannot be compensated later due to the short and intensive Finnish growing season (Peltonen-Sainio et al. 2009b, 2009c, 2011).

Hail is a form of solid precipitation that consists of balls or irregular lumps of ice. Hail is seldom observed in a traditional meteorological network, because precipitation in the form of hail is a small-scale event that lasts only for a short time. The methodology to detect hail from radar measurement data is gradually improving. In this study we could not separate hail from the total precipitation.

Minimum air temperatures

Minimum air temperatures below zero are possible throughout the growing season in Finland, although they are most likely and frequently experienced at the beginning and at the end of the growing season (Figure 3). Below-zero air temperatures during the growing season, i.e. night frosts (FROST), have caused the worst famines over the centuries in Finland. Up until the 1950s, night frost caused frequent yield losses every fifth year on average, but it has occurred less frequently since then (Mukula and Rantanen 1987). In cereals such as wheat, night frost seldom causes total crop failure when it occurs early or late in the growing season. Wheat can survive very low temperatures, even when repeated night after night. However, early summer night frost retards growth for as much as couple of weeks, which causes yield penalties in our short growing season (Peltonen-Sainio et al. 2009b). Total crop failures are only caused by night frost when it occurs at anthesis and thereby destroys the grain primordia, which have a high water content prior to the onset of grain-filling. Hence, the timing of night frost is very important.

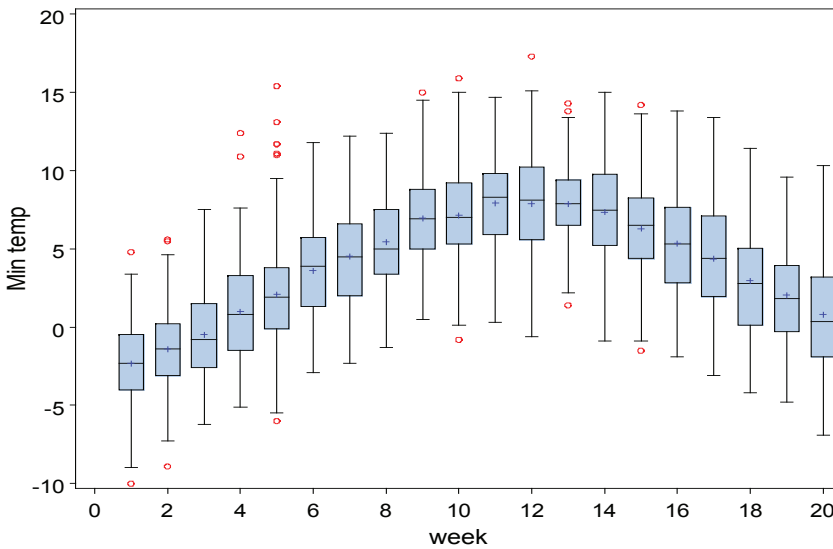*Pietola, K. et al. Predicting yield of spring wheat by weather*

Fig. 3. The weekly minimum temperatures in 1970-2008. The six sample regions are pooled. Tukey box: see Fig. 1 for details.

Yield data

Over the years, spring wheat yields have varied within a large range (Figure 4). The mean yield was within the full sample 3,850 kg ha[-1], and the yield varied from the lowest 1,531 to the highest 6,198 kg ha[-1].

Some differences in yields occurred between sample regions (Table 1). The regional yields reflect climate conditions, which are more favourable in southern parts of Finland (e.g. Mietoinen, Pälkäne and Jokioinen) than in northern parts of the country (e.g. Ylistaro).
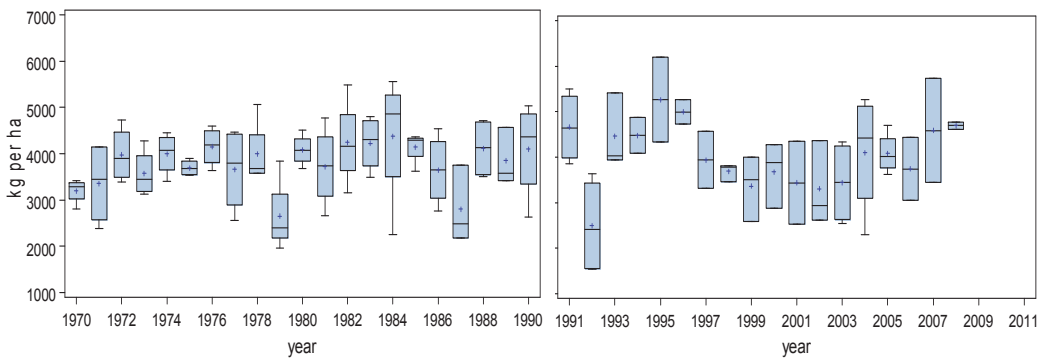

Fig. 4. Spring wheat yields in 1970-2008. The sample regions are pooled. Tukey box: see Fig. 1 for details (MTT Official Variety Trials)

Table 1. Spring wheat yields (kg ha^{-1},) in the sample regions pooled across years (1970-2008). Data from the MTT Official Variety Trials 1970–2008. The yield data are normalized to fixed fertilization levels.

Region	N	Mean	SD	Minimum	Maximum
Jokioinen	28	3,863	944	1,581	5,734
Mietoinen	36	3,713	722	2,414	5,273
Pälkäne	36	4,201	773	2,383	6,197
Ylistaro	34	3,612	988	1,531	5,499

Model and Methods

Indemnity payment function for the index insurance contract

Following (Barnet et al. 2005) we define the indemnity payment (n) for the weather index insurance contract as:

$\tilde{n}_t^w(\tilde{w}_t) = \max(\overline{w}^c - \tilde{w}_t, 0)$, for shortage (2a)

$\tilde{n}_t^w(\tilde{w}_t) = \max(\tilde{w}_t - \overline{w}^c, 0)$, for excessiveness (2b)

where

$\tilde{w}_t$ = a function of the stochastic weather indices at time t

$\overline{w}^c$ = the vector of critical values that trigger the payment

The indemnity payment depends on the realization of the stochastic weather index values through function $\tilde{w}_t$, and the corresponding critical values for these indices $\overline{w}^c$.

Estimating equations for the indemnity payments

To get well defined, indemnity payments over the relevant regimes we define the underlying function ($\tilde{w}_t$) piece-wise linear in two dimensions with respect to the three above described weather measurements (m): Growing Degree Days (GDD), precipitation (RAIN) and night frost (FROST). The first dimension is with respect to time so that the marginal yield effect of each weather measure is allowed to differ between time regimes. The location and the duration of these time regimes along the cropping season are tested. In this case, we regress the yield (yt) on the index function $\tilde{W}_t$ and the estimating equation is[6]:

---

6 Without losses in generality, we have normalized the price of wheat to one and the marginal yield effect

($\partial y(m) / \partial m_i$) equals to the tick price of the contract.

Pietola, K. et al. Predicting yield of spring wheat by weather

$$y_t(m) = \alpha + \beta^\ell D^\ell + \sum_{\tau=1}^{T} \phi^\tau GDD_t^\tau + \sum_{\kappa=1}^{K} \varphi^\kappa RAIN_t^\kappa + \sum_{\gamma=1}^{\Gamma} \theta^\gamma FROST_t^\gamma + \varepsilon_t \tag{3}$$

where $\alpha, \beta, \phi, \varphi$ and θ

are parameters and ε_t is an error. The subscript t indices year and the index function is defined at annual frequency to match the annual yield data. The superscript ℓ refers to location and the dummy variable (D) is used to allow different conditional means for different locations. The indices $\tau=1,...T$; $\kappa=1,...K$ and $\gamma=1,...\Gamma$ distinguish between different regimes in the piece-wise linear specification.

The second dimension for piece-wise linearity is with respect to the values of the weather measures. This way we allow for asymmetry in marginal yield effects with respect to the values of the weather measures. For GDD we distinguish three different regimes WARM, NORMAL and COLD. WARM and COLD regimes are identified by dummy variables DWARM and DCOLD that receive value one if the summer is warm or cold and value zero otherwise. The middle regime (NORMAL) is observed if $D^{WARM} = D^{COLD} = 0$. For RAIN we consider DRY, NORMAL and WET regimes with the corresponding dummy variables D^{DRY} and D^{WET}. The underlying idea is that the marginal yield effect for RAIN, for example, is expected to differ significantly if the growing season is DRY, NORMAL or WET. The regimes are optimized by a grid search method so that the mean square error between the yield and the weather measurements is minimized[7]. The asymmetric estimating equation is:

Our estimation approach is to start projecting the yields conditional on each weather measurement separately. Estimating these first-stage "partial" and incomplete specifications is consistent under the null-hypothesis that the weather does not significantly imply the yields. For those weather indices that turn out significant, we later estimate their effects jointly in the second stage. These second-stage estimations provide more efficient estimates. We may also expect that the weather indices have significant joint effects similar to Leontief technologies (e.g. Chambers 1988), because the different weather attributes are more likely to complement rather than substitute for each other[8].

In the first stage, we use weekly data separately regressed on the GDD, FROST, and the RAIN. These weekly data are used to test on how best aggregate the weather indices further across time so that simple indices can be constructed and the degrees of freedom allow estimation of the yield response to all relevant weather indices jointly. To maintain sufficient over-identification restrictions in estimation, the site-specific data are pooled in most of the specifications, and the error correlation induced by unobservable time-invariant site-specific effects is factored out using dummy variables (Wooldridge 2002, p. 133).

$$
\begin{aligned}
y_t(m) = {}& \alpha + \beta^\ell D^\ell \\
& + D^{COLD} \sum_{\tau=1}^{T^{COLD}-1} \phi^\tau GDD_t^\tau + \sum_{\tau=T^{COLD}}^{T^{WARM}-1} \phi^\tau GDD_t^\tau + D^{WARM} \sum_{\tau=T^{WARM}}^{T} \phi^\tau GDD_t^\tau \\
& + D^{DRY} \sum_{\kappa=1}^{K^{DRY}-1} \varphi^\kappa RAIN_t^\kappa + \sum_{\kappa=K^{DRY}}^{K^{WET-1}} \varphi^\kappa RAIN_t^\kappa + D^{WET} \sum_{\kappa=K^{WET}}^{K} \varphi^\kappa RAIN_t^\kappa \\
& + \sum_{\gamma=1}^{\Gamma} \theta^\gamma FROST_t^\gamma + \varepsilon_t
\end{aligned}
\tag{4}
$$

---

7 The grid search is to give alternative locations and durations for the regimes and then select the one with the best fit.

8 In Leontief technology a production factor cannot substitute for another production factor. Here it would imply that rainfall, for example, cannot substitute for the shortage of GDD.

Results

Growing degree days (GDD)

Our hypothesis is that, within the range of sample variation, GDD should have a positive effect on the yield in the late growing season (July−August), but in early growing season (May−June) the effect may also be negative or insignificant, since moderate GDD values in early summer may increase the yield potential, which might be realized later in the growing season.

When the models are estimated separately for each experimental site using weekly GDD values, they explain 62−74% (R^2) of yield variations. Nevertheless, in these site-specific models we found most of the parameter estimates to be insignificant and alternating their signs, suggesting an inconsistent specification and lack of over-identification restrictions required for obtaining consistent, well identified parameters. In general, it is possible that unfavourable conditions causing yield penalties at a relatively early stage of the growing season are partially compensated for by favourable growing conditions later on. However, especially in northern growing conditions, such compensation is marginal due to the exceptionally short growing season and long days that enhance development of the wheat crop (Peltonen-Sainio et al. 2007, 2009b, 2009c). This can actually be considered to give a good premise for relating weather indices to yield losses.

When the site-specific data are pooled, the weekly GDD explain 14% of the yield deviations.[9] Nevertheless, the site-specific dummy variables were found to be insignificant, suggesting that once the GDD are controlled for, the location does not, as such, have a significant effect on the yield of wheat within the typical wheat production areas. Again, most of the parameter estimates on the weekly GDD were insignificant, and their signs alternated inconsistently. Overall, the pattern of the point estimates indicates that, within the sample variation, an increase in the cumula-

tive temperatures significantly implies the yields in early summer (June) and late summer (August), but not significantly in spring (May) or immediately after midsummer (July).

Aggregating the GDD data into two-week periods improves the consistency of the estimates and more clearly suggests than the results above that the GDD significantly implies yields in early summer and after midsummer (Figure 5). The impact is first slightly negative in the early growth stages when yields are mostly determined, and then increases to become positive towards later summer. However, later on in the autumn and at harvest, the effects of GDD decline and once again become insignificant. This result is consistent with our hypothesis and earlier experience, so that a moderate GDD in early summer supports the formation of the yield potential that may or may not be realized, depending on the GDD and other conditions later on in the summer. In September it is likely that high temperatures associate with low precipitation, together promoting the ripening processes and keeping plant stands unlodged, although they occur too late to enhance yields. Therefore, we drop the last observations from the estimating equations.
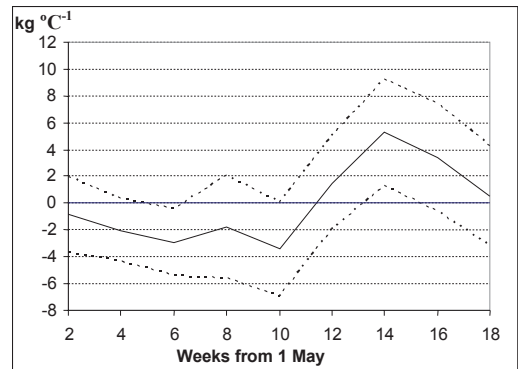

Fig. 5. The marginal yield effect of the cumulative degree days (GDD) for two-week periods. Thick line: the estimate , thin lines: the estimate plus/minus its standard errors. Estimated in Equation 3, imposing $\varphi^\kappa = \theta^\gamma = 0$ for all κ and γ.

9 The regional dummy variables are insignificant and alone explain 0.8% of the yield variation.

Pietola, K. et al. Predicting yield of spring wheat by weather

When the data on GDD are further aggregated into four-week cumulative values and the last (fifth) insignificant GDD variable is dropped, the projection explains 6.3% (R^2) of the yield variation. In this case, the data also support the main idea that the accumulation of temperature most clearly influences the yield in early summer (June) and in late summer (August) (Figure 6). The second four-week period, which corresponds to June, has a significant negative effect on the yield. The fourth four-week period GDD, which corresponds to August, has the largest and the only significant positive effect on the yield. Thus, around the sample means, GDD has not significantly influenced the yields in spring (May) or in midsummer (July). An intercept and the fourth GDD (August) variable alone explain about 4% of the yield deviations. Within our sample, these August temperatures more strongly influence yields than the total GDD for the whole growing season.
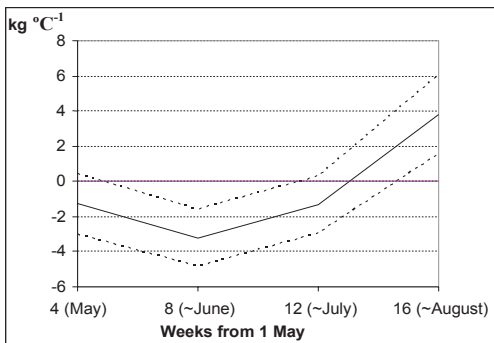

Fig. 6. The marginal yield effect of the cumulative degree days (GDD) for four-week periods. Thick line: the estimate; thin lines: the estimate plus/minus its standard error. Estimated in Equation 3, imposing $\varphi^\kappa = \theta^\gamma = 0$ for all κ and γ.

We might expect that under Finnish weather conditions the cold periods and shortage of GDD would delay crop growth and decrease the yield. We therefore tested for asymmetry in the GDD by defining distinct regimes for colder than average, and warmer than average temperatures respectively. The thresholds "mean minus 10 ºC" for the cold regime and "mean plus 25 ºC" for the warm

regime in the four-week GDD aggregates provided the best fit for the sample. Nevertheless, the data identify significant asymmetries only in the third and fourth four-week periods, suggesting that at lower than average temperatures the marginal yield effect of one GDD is higher than at the mean or at higher than average temperatures (Table 3, Figure 7). In the cold regime, the marginal value of one GDD point increases for the third four-week period from 17 to 19 kg.[10]
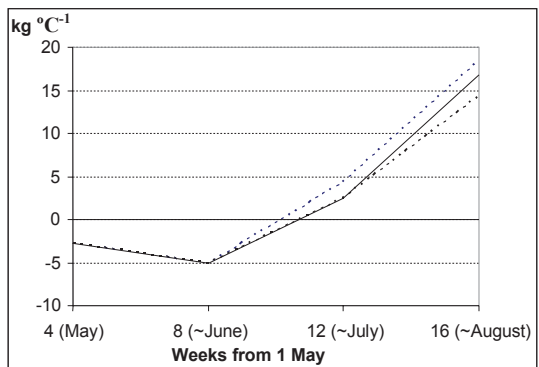

Fig. 7. The marginal yield effect of from asymmetric projection of yield on GDD for four-week periods. Thick line: middle regime, uppermost thin dotted line: COLD regime, lower thin dotted line: WARM regime. Estimated in Equation 5, imposing $\varphi^\kappa = \theta^\gamma = 0$ for all κ and γ.

Minimum air temperatures: frosts

The night frost is defined as a day when the daily minimum air temperature falls below a threshold. Then the frost variable (FROST) equals that minimum temperature, and otherwise it equals zero. The threshold was first defined at 0, −1 or −2 ºC. However, the data for the latter two threshold values turned out sparse and the parameters were not so well identified. We therefore report only the results for the model using 0 ºC as the threshold.

The data suggest that frost is the most critical factor in the middle of July. The most damaging and significant frost effects are identified for the

---

10 Our approach of allowing and testing for asymmetry and the threshold effects is commonly adopted in the econometric analysis of time series for price movements (see. e.g. Serra et al. 2006; Jalonoja and Pietola 2004)

Table 3. The marginal yield effects (kg ºC^{-1}) of asymmetric projection of yield on GDD for four-week periods. Estimated in Equation 5 with the restriction $\varphi^{\kappa}=\theta^{\gamma}=0$ for all κ and γ.

Variable	Weeks[1]	Parameter	Std error
Intercept		-4,470	1,500
Location: Mietoinen		9.06	210
Location: Palkane		129	210
Location: Ylistaro		-158	219
GDD	1-4	-2.81	1.75
GDD	5-8	-5.09	1.63
GDD	9-12	2.41	2.23
GDD	13-16	16.8	4.68
GDD, COLD[2]	9-12	1.86	0.75
GDD, COLD[2]	13-16	1.75	0.934
GDD, WARM[2]	13-16	-2.41	0.776

[1]Counting the weeks starting at 1 May.
[2]Thresholds: average GDD of the period -10 ºC for COLD, and average GDD of the period + 25 ºC for WARM.

two-week period in mid-July (bi-weekly aggregate ending at week number 12, Figure 8). The point estimate for the losses due to frost in mid-July is 2,400 kg, with a standard error of 1,500 kg. Thereafter, and towards the harvest, the adverse yield effect rapidly declines and becomes insignificant. Early in the growing season (May) or at harvest (September), frosts do not result in significant crop damages. In June, some signals of the adverse effects of frost emerge, but the data are not informative enough to identify these effects.

If only the statistically significant parameters are included in the estimating equation, the magnitude of the yield loss due to mid-July frost remains at 2,100 kg. If we estimate the frost effect using a dummy variable for the minimum temperatures below zero, the point estimate for the mid-July frost effect is 1,200 kg, with a standard deviation of 880 kg. Nevertheless, the earlier specification performs better than this version.

Rainfall

Our hypothesis is that within the range of variation observed in our sample, rainfall has both negative and positive yield effects, depending on the timing and amount. With respect to timing, we identify
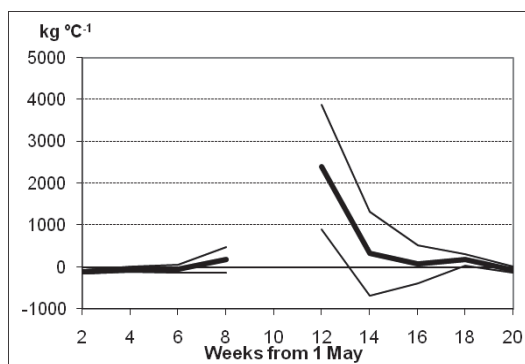
Fig. 8. The marginal yield effect in regressing wheat yield deviations on the night frost (FROST) in each two-week period. Thick line: the estimate, thin lines: the estimate plus/minus its standard error. The middle part is not identified. Estimated in Equation 3, imposing $\phi^{\tau}=\varphi^{\kappa}=0$ *for all* τ *and* κ.

three distinct regimes. The first period is defined for spring at the beginning of May, when rainfall is expected to reduce the yield potential by either delaying sowing or by causing crust formation. The second regime is in late May and June, when the deficiency of rainfall and the risk of drought are greatest and rainfall (if not excessive) would have a positive yield effect. Thereafter, towards the end of July and August (3rd time period), the positive effects of rainfall gradually decline, because rain comes too late for the crops and begins to adversely

affect the harvest. Besides all these impacts within the three regimes, extremely intensive excess rainfall adversely affects yields and may result in crop damage throughout the growing season, from sowing to harvest.

We have two separate measures for rainfall. The first is the weekly sum of rainfall and, according to the patterns observed in the data, it is then aggregated over different durations within the three regimes described above. The second measure focuses on excess rainfall. It is defined as the frequency of rains that exceed a minimum threshold within two days. The threshold levels are set at 40 and 60 mm.[11] However, these frequencies were found to be sparse and therefore difficult to link to yield deviations, as was also shown when using daily precipitation values exceeding 10 mm or 20 mm in Peltonen-Sainio et al. (2009a). Hence, these frequency data are not informative enough to identify the adverse yield effects of excessive rainfall.

In the linear and symmetric projections, the weekly rainfall explains 23.4% of the deviations in yield when all four locations are pooled and the site-specific differences in the mean yields are controlled for with dummy variables. The dummy variables were also insignificant in this specification, suggesting that once the rainfall measures are controlled for, the location does not have a significant effect on the yield of wheat within the typical wheat production areas. Similarly to GDD measures, the parameter estimates for the weekly rainfall measures alternated their signs, suggesting that a weekly aggregation interval is too short for consistently identifying rainfall effects in the data. We therefore aggregated the rainfall measures using bi-weekly intervals. Then the rainfall implies more consistently and significantly the yield over the growing season.

The results suggest that before and immediately after sowing in early May, rainfall reduces the yield, which may also imply problems in sowing in good time. Thereafter, rainfall in midsummer

is predicted to increase the yield, and then later on the positive yield effect starts to diminish and becomes insignificant towards the harvest. Immediately before and at harvest, rainfall reduces the yield (Figure 9). This specification explains 17.0% of the total yield variation within the sample.
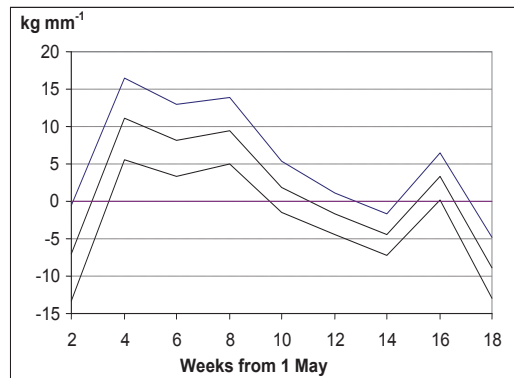

Fig. 9. The marginal yield effect in regressing the wheat yield deviations on the cumulative rainfall per two-week period. Thick line: the estimate, thin lines: the estimate plus/minus its standard error. Estimated in Equation 3, imposing $\phi^\tau = \theta^\gamma = 0$ *for all* τ *and* γ.

Aggregating the rainfall measures further across time improves the consistency of the estimates and highlights the three regimes with distinct yield impacts (Figure 10). This specification has the power to explain 14.6% of yield variation within the sample. The marginal product of 30 mm rainfall, for example, is estimated in early May at −200 kg, and then it increases to 300 kg. At the beginning of July the marginal product starts to decline, being zero in August and becoming −200 kg in early September.

We could expect that the effect of rainfall on the yield is asymmetric and has thresholds, especially in early and mid-summer, when drought is a concern and a deficiency of rainfall restricts crop growth. Thus, below a certain threshold, rainfall could have a larger positive impact on the yield than above the threshold. Similarly, the yield impact may become smaller or negative when rains are excessive, especially when the harvest is approaching. Once the lower (DRY) and upper

11 We also computed the frequency of rains exceeding 80 or 100 mm within two days. The sample did not include any two days of rain exceeding 100 mm, and the number of nonzero frequencies for rainfall of more than 80 mm was too small for identification.

(WET) regimes are allowed to differ from the middle regime, the middle regime may be expected to be characterized as a regime of "inaction" in which the effects of rainfall are smaller than in the lower and upper tails. We therefore further augment the estimating equation with asymmetric yield effects
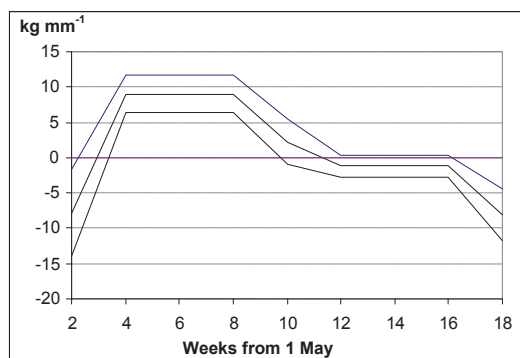

Fig. 10. The marginal yield effect in regressing the wheat yield deviations on the cumulative rainfall. The horizontal straight lines indicate regimes where parameters are imposed equally. Thick line: the estimate, thin lines: the estimate plus/minus its standard error. Estimated in Equation 3, imposing $\phi^\tau = \theta^\gamma = 0$ *for all τ and γ.*

with respect to the amount of rainfall. The yield impact of rainfall is separated into three regimes: a dry regime, middle regime and wet regime.

The data were informative enough to identify the distinct effects of drought in early summer and excess rainfall in late summer and towards the harvest (Table 4, Figure 11). The lower threshold of 30 mm per two weeks and higher threshold of 50 mm per two weeks provided the best sample fit. At the beginning of the growing season and at harvest, the yield response is nevertheless symmetric and negative so that all rainfall reduces the yield. Around midsummer, when drought is common concern, the marginal yield effect of rainfall within the dry regime is estimated at 15.8 kg, suggesting that 30 mm of rainfall within the dry regime increases the yield by almost 500 kg. Within the wet regime, the marginal yield effect of rainfall decreases earlier and faster to a negative value as compared to the other regimes. Within the wet regime the marginal yield effect becomes negative already by the end of July.

Table 4. The parameter estimates and their standard errors in the asymmetric model for the wheat yield deviations conditional on the cumulative rainfall (RAIN). Estimated in Equation 5, imposing $\phi^\tau = \theta^\gamma = 0$ *for all τ and γ.*

Variable	Weeks[1]	Parameter	Std error
Intercept		-683	405
Mietoinen		-26.5	205
Palkane		-126	204
Ylistaro		-345	210
RAIN	1-2	-6.63	5.87
RAIN	3-10	9.73	2.47
RAIN	11-16	1.44	1.91
RAIN	17-18	-7.02	3.58
RAIN, DRY[2]	3-8	6.04	2.83
RAIN, WET[2]	9-10	-7.60	3.00
RAIN, WET[2]	11-14	-3.79	1.68

[1]Counting the weeks starting from 1 May.
[2]DRY regime: rainfall less than 30 mm per two weeks; WET regime: rainfall exceeds 50 mm per two weeks.
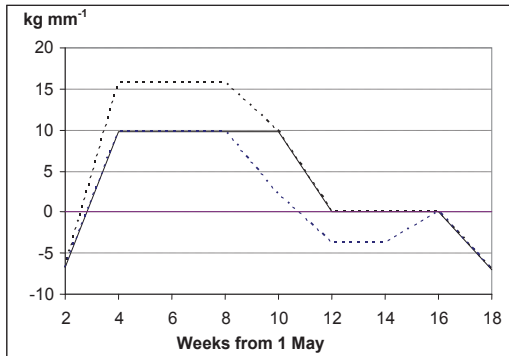
Fig. 11. The marginal yield effect in the asymmetric model for wheat yield deviations conditional on the cumulative rainfall. Thick line: middle regime, uppermost dotted thin line: DRY regime (rainfall less than 30mm per two weeks), lower dotted line: WET regime (rainfall exceeds 50 mm per two weeks). Estimated in Equation 5, imposing $\phi^\tau = \theta^\gamma = 0$ *for all τ and γ*.

Joint effects

The joint effects of GDD, FROST and RAIN of the yield are estimated by regressing the yield on all variables jointly, so that each of them enter the equations similarly to the above-described with asymmetric effects. This model is estimated in equation 5 and it explains 38% of the yield variation. The signs of the parameter estimates remain unchanged for all weather indices in the joint model as compared to the separate models, but the magnitudes of some of the parameters change (Table 5).

We then tested for heteroskedasticity by first regressing the squared error of the ordinary least squares (OLS) specification "All" in Table 3 on an intercept, fitted values and the squared fitted values. This testing approach combines the features of Breusch-Pagan and White tests, but requires only two degrees of freedom. For testing purposes, the approach is valid under the null hypothesis of homoskedasticity (Wooldridge 2002 p. 127). This auxiliary regression suggests that the OLS model is heteroskedastic, implying that the

estimates are consistent but not efficient. The results reveal thatyield volatility decreases with the yield. In other words, low yield regimes are more volatile than high yield regimes, suggesting that the growing conditions causing crop damage are likely to have heterogeneous implications and that they therefore also increase yield volatility.

We further tested for heteroskedasticity by regressing all explanatory weather events used in the specification "All" on its squared error. This auxiliary specification suggests that dry and warm early summer periods in June significantly increase yield volatility. The yield volatility is higher at Mietoinen and Palkane than at the other locations.

Correcting the model for homoskedasticity, we took the fit determined by the significant parameters in this (latter) auxiliary equation by imposing all other parameters at zero, and computed the weight variable for the feasible generalized least squares (FGLS) by taking a square root of this fit. The parameter estimates for the re-estimated FGLS specification are similar to those in the standard OLS specification (Table 6). The results therefore suggest that the standard OLS estimates, which are known to be consistent but not efficient if the error is heteroskedastic, are robust to alternative corrections for heteroskedasticity.

Conclusions

Our results suggest that growing degree days (GDD), rainfall (RAIN) and night frosts (FROST) significantly influence the yield of wheat in Finland. When the weather measures are aggregated to obtain well-identified and consistent estimates as well as simple indices, as required by insurance contracts and tradable weather derivatives, they jointly explain about 38% of the yield variation. Thus, using simple weather event-based indices as in our model, about 38% of wheat grower yield risk could be insured at best, with the remaining 62% being left as uninsured basis risk.

Within the sample, rainfall contributes the most to the yield and alone explains 23% of the yield

Table 5. The parameter estimates and their standard errors in the asymmetric models regressed separately on GDD and RAIN, and on all weather events jointly. Estimated in Equation 5.

Model variable	Weeks starting from May 1	GDD		RAIN		All	
		Parameter	Std. Error	Parameter	Std. Error	Parameter	Std. Error
Intercept		-4,470	1,500	-683	405	-5,160	1,420
Mietoinen		9.060	210	-26.5	205	32.5	195
Palkane		129	210	-126	204	-59.0	194
Ylistaro		-158	219	-345	210	-309	204
GDD	1-4	-2.81	1.75			-2.39	1.67
GDD	5-8	-5.09	1.63			-3.03	1.54
GDD	9-12	2.41	2.23			1.19	2.08
GDD	13-16	16.8	4.68			16.3	4.42
GDD, COLD[1]	9-12	1.86	0.748			1.92	0.69
GDD, COLD[1]	13-16	1.75	0.934			1.97	0.88
GDD, WARM[1]	13-16	-2.41	0.776			-2.37	0.72
RAIN	1-2			-6.63	5.87	-9.47	5.69
RAIN	3-10			9.73	2.47	9.30	2.39
RAIN	11-16			1.44	1.91	1.49	1.92
RAIN	17			-7.02	3.58	-7.82	3.43
RAIN, DRY[2]	3-8			6.04	2.83	5.55	2.66
RAIN, WET[2]	9-10			-7.60	3.00	-7.06	2.91
RAIN, WET[2]	11-14			-3.79	1.68	-2.70	1.60
FROST	11-12					3,120	1,270
R^2 %							

[1] COLD: cold regime; WARM: warm regime.
[2] DRY: dry regime; WET: wet regime.

Pietola, K. et al. Predicting yield of spring wheat by weather

Table 6. The parameter estimates and their standard errors in OLS and FGLS models. Estimated in Equation 5.

Model variable	Weeks starting from May 1	OLS (All)		FGLS (All)		FGLS (Restricted)	
		Parameter	Std. Error	Parameter	Std. Error	Parameter	Std. Error
Intercept		-5,160	1,420	-4,780	1,450	-4,870	1,37
Mietoinen		32.5	195	-40.6	197		
Palkane		-59.0	194	29.2	198		
Ylistaro		-309	204	-261.	213	-256	162
GDD	1-4	-2.39	1.67	-2.21	1.65	-2.19	1.63
GDD	5-8	-3.03	1.54	-3.45	1.56	-3.44	1.52
GDD	9-12	1.19	2.08	-0.26	2.05		
GDD	13-16	16.3	4.42	16.7	4.61	16.7	4.51
GDD, COLD[1]	9-12	1.92	0.69	1.48	0.67	1.54	0.49
GDD, COLD[1]	13-16	1.97	0.88	1.78	0.89	1.78	0.87
GDD, WARM[1]	13-16	-2.37	0.72	-2.40	0.71	-2.39	0.70
RAIN	1-2	-9.47	5.69	-11.3	5.69	-11.4	5.54
RAIN	3-10	9.30	2.39	9.69	2.42	9.88	2.32
RAIN	11-16	1.49	1.92	1.52	1.91	1.51	1.85
RAIN	17	-7.82	3.43	-6.08	3.45	-6.12	3.38
RAIN, DRY[2]	3-8	5.55	2.66	5.87	2.64	5.86	2.53
RAIN, WET[2]	9-10	-7.06	2.91	-6.56	2.92	-6.64	2.84
RAIN, WET[2]	11-14	-2.70	1.60	-2.14	1.55	-2.13	1.52
FROST	11-12	3,120	1,270	3,100	1,310	3,090	1,300
R^2 % [3]		37.9		39.4		39.3	

[1] COLD: cold regime; WARM: warm regime.

[2] DRY: dry regime; WET: wet regime.

[3] R^2 is not fully comparable between the OLS and FGLS specifications, since FGLS is rescaled.

variation, whereas the corresponding explanatory power for GDD was estimated at 16%. Thus, when the production site and the weather measurement points are at the same location, rainfall-based measures seem to have larger potential for hedging against yield risks, but it is known that rainfall has more spatial variation than temperature-based GDD. Therefore, when the distance from the weather station to the production site increases, the relative efficiency of the rainfall- and temperature-based measures may become ambiguous and reversed. These spatial questions require different analysis and are addressed in Myyrä et. al (2011).

The estimates are consistent with our expectations, in terms of that each of these three measures has qualitatively and quantitatively different yield impacts depending on period. The effect of GDD is initially negligible, as a mild early growing season may even increase the yield potential. The marginal yield effect of GDD increases towards August and peaks within the period from late July to mid-August (13-16 weeks from the beginning of May). In this period the marginal yield effect of a GDD point is estimated at 17 kg ºC^{-1}. Within the cold regime, i.e. when the accumulation of GDD

is a particular concern, the corresponding marginal yield effect is estimated even higher (19 kg ºC^{-1}).

Another temperature-related index, night frost (FROST), is estimated to be the most critical, with the most severe crop damage in the middle of July. The marginal yield effect of night frost in the middle of July was estimated to be as high as 2,100-2,400 kg of wheat, depending on the specification of the frost variable. In other words, if a night frost hits in mid-July, it is expected to result in yield losses of between 2,100 and 2,400 kg per hectare. At the beginning and towards the end of the growing season, the damage caused by night frost is estimated to be negligible.

Rainfall has three separate regimes with respect to time. First, at the beginning of May, the approximate sowing time in Southern Finland, any significant rainfall is estimated to reduce the yields, since it delays sowing and excessive rains may also result in crust formation. However, in June, when drought typically influences crop yields, the marginal yield effect of rainfall peaks at a significant positive value of 10 kg mm^{-1}. Thereafter, in July and August, the impact of rain gradually declines with the passage of time and turns negative at the end of August. Rain in late July and thereafter, comes too late to improve the growth and begins to adversely affect both standing yields and their harvest.

With respect to rainfall, the data are not informative enough to identify the negative yield effects of excessive rainfall, when the rainfall is measured as the frequency of excessive rain defined as, for example, at least 40 mm within a two-day period. The data are nevertheless informative enough to identify asymmetric rainfall effects early and late in the growing season. In the early season the dry regime differs from the other regimes, and in late summer the wet regime differs from the others. In June, when a shortage of water is a particular concern, the marginal yield effect of one millimetre of rain is estimated within the dry regime at 16 kg. Within the wet regime, on the other hand, the marginal yield effect of rain decreases earlier and faster than in other regimes to negative values in July and August, towards the harvest. The marginal yield effect is estimated at the lowest and negative

value (-7 kg mm^{-1}) at the end of the growing season in all regimes.

In addition to the above-described effects on the expected yields, the data indicate that weather events have heterogeneous yield implications so that, within the range of sample variation, the yield volatility increases with adverse weather events causing crop damage. In particular, dry early and midsummer periods in June increase yield volatility and reduce yields.[12]

Our results have several important implications for the design of simple and tractable (efficient) weather index-based insurance contracts. First, the weather events triggering the indemnity payments should be focused on certain critical periods and regimes. When protecting against yield losses resulting from a shortage of GDD, the period from late July to mid-August is the most critical. When the shortage of GDD accumulation is of particular concern (the cold regime), the marginal yield effect of one GDD unit is estimated to be as high as 19 kg of wheat. If, for instance, the daily temperature is decreased by one degree over a month, the GDD is reduced by 31 degree points and the resulting yield loss per hectare is estimated at 590 kg.

Second, when protecting the grain grower against yield losses caused by night frosts, mid-July is the most critical time. The likelihood for a night frost in July is low, but if a frost hits it results in yield losses of more than 2,100 kg per hectare.

Third, when protecting grain grower against yield losses from a shortage of rainfall, June is the critical month, when the marginal yield effect of one millimetre of rain is estimated at 16 kg. Thus, 30 mm of rainfall has a marginal product of almost 500 kg. However, if excessive rainfall is a concern, the contract should focus on the beginning and end of the growing season.

When the distance between the location of weather measurement and the standing crop increases, the spatial correlation between the weather events becomes an issue (e.g. Myyrä et al. 2011). The higher the spatial correlation is the more prom-

---

12 The result holds only locally within the range of sample variation. The asymptotic property should be that when a weather event is so severe that the yield approaches zero, the volatility also approaches zero.

A G R I C U L T U R A L   A N D   F O O D   S C I E N C E

Pietola, K. et al. Predicting yield of spring wheat by weather

ising the indices would be in establishing liquid market regimes to trade these contracts. But if the correlation is low (as in the case of rain and hailing) the weather index measurement has to be located at a very close to the standing crops to be relevant and efficient in hedging. In this case the contracts have to be tailored to local measurement points and can perhaps be sold at the best over the counter. These spatial questions are, nevertheless, left as a topic for another research. This paper addresses only the correlations between the weather indices and yield when the weather measurement point and the standing crops are at the same location.

Finally, complying with the WTO and CAP regulations would be a challenge for a subsidized index based contract, if these regulations require that only yield losses of more than 30% can be subsidized and receiving the indemnity payments require that the losses have to be observed on the farm. Since index-based contracts are only indirect measures of the true yield losses, proving yield losses of more than 30% may ultimately be challenged by the heterogeneity of the weather effects, as our results suggest that not only are expected yields reduced, but the yield volatility is also increased when adverse weather shocks are realized.

References

Barnett B. J. & Mahul O. 2007. Weather Index Insurance for Agriculture and Rural Areas in Lower-Income Countries. *American Journal of Agricultural Economics* 89: 1241-1247.

Barnett B.J., Black J.R., Hu Y., & Skees J.R. 2005. Is Area Yield Insurance Competitive with Farm Yield Insurance? *Journal of Agricultural and Resource Economics* 30: 285-301.

Chambers, R. G. 1988. *Applied Production Analysis, A Dual Approach*. Cambridge University Press: New York. 331 p.

CME Group 2010. *Chicago Board of Trade and NYMEX Company*. Cited 24 Nov 2008. Available on the Internet: http:// www.cmegroup.com/trading/weather

Jalonoja, K. & Pietola K. 2004. "Spatial Integration between Finnish and Dutch Potato Markets". *Food Economics* Vol 1: 12-20.

Kangas A., Laine A., Niskanen M., Salo Y., Vuorinen M., Jauhiainen L. & Nikander H. 2009. Virallisten lajikekokeiden tulokset 2002-2009 . *MTT Kasvu* 6 : 1-184.

Karuaihe, R.N., Wang, H.H. & Young, D.L. 2008. *Farmers' Demand for Weather-Based Crop Insurance Contracts: The Case of Corn in South Africa.* Working Paper. Purdue University.

Koundouri, P., Laukkanen, M., Myyrä, S. & Nauges, C. 2009. The effects of EU agricultural policy changes on farmers´ risk attitudes. *European Review of Agricultural Economics* 36: 53-77.

Miranda, M.& Vedenov, D.V. 2001. Innovations in Agricultural and Natural Disaster Insurance. *American Journal of Agricultural Economics* 83: 650-655.

Mukula J. and Rantanen O. (1987). Climate risk to the yield and quality of the field crops in Finland. *Annales Agriculturae Fenniae* 26:1-18.

Myyrä, S., Pietola, K. & Jauhiainen, L. 2011. Systemic Yield Risk and Spatial Index Correlations: Relevant Market Area for Index Based Contracts. *Food Economics* 8: 114-125. Available online. DOI:10.1080/16507541.2011.606647.

Peltonen-Sainio, P., Kangas, A., Salo, Y. & Jauhiainen, L. 2007. Grain number dominates grain weight in cereal yield determination: evidence basing on 30 years' multi-location trials. *Field Crops Research* 100: 179-188.

Peltonen-Sainio, P., Jauhiainen, L. & Hakala, K. 2009a. Are there indications of climate change induced increases in variability of major field crops in the northernmost European conditions? *Agricultural and Food Science* 18: 206-226.

Peltonen-Sainio, P., Rajala, A., Känkänen, H. & Hakala, K. 2009b. Improving farming systems in northern European conditions. In: Victor O. Sadras and Daniel Calderini (eds.). *Crop Physiology: Applications for Genetic Improvement and Agronomy*. Amsterdam: Elsevier. p. 71-97.

Peltonen-Sainio, P., Jauhiainen, L., Rajala, A. and Muurinen, S. 2009c. Tiller traits of spring cereals in tiller-depressing long day conditions. *Field Crops Research* 113: 82-89.

Peltonen-Sainio, P., Jauhiainen, L. & Hakala, K. 2011. Crop responses to temperature and precipitation according to long-term multi-location trials at high-latitude conditions. *Journal of Agricultural Science* 149: 49-62.

Serra, T., Gil, J.M. & Goodwin, B. 2006. Local Polynomial Fitting: Price Transmission in EU Pork Markets. *European Review of Agricultural Economics* 33: 415-436.

Xu, W., Filler, G., Odening, M. & Okhin, O. 2009. Selected paper prepared for presentation at the 2009 AAEA & ACCI Joint Annual Meeting in Milwaukee, Wisconsin, July 26 −28.

Van Asseldonk, M. & Oude Lansink A. 2003. Weather-Based Index Insurance to Hedge Temperature Exposure of Greenhouse Horticultural Firms. In: Wesseler, J.,Weikard, H-P. & Weaver, R.D. (eds.). *Risk and Uncertainty in Environmental and Natural Resource Economics*. Northampton, USA: Edward Elgar Publishing. p. 235-49.

Wooldrige, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. London, England: The MIT Press. 752 p.