Timothy Tyree
Reed College
Econometrics 312 Final Project
Jeff Parker
May 8, 2014
Word Count = 6,437

Analysis of the Effect of Elevation on Glacial Mass Balance

I. Introduction

This study uses 274 stake measurements of the mass balance of the Gulkana Glacier in 1966-1968 to compare polynomial functional forms of mass balance profiles. *Mass balance* is the change in elevation at a point over the course of a year. *Mass balance profiles* are models that describe the mass balance over the entirety of a glacier as a function of elevation above sea level. This study concludes that a quartic function of elevation best models mass balance. Classical Glaciology predicts a cubic relationship between elevation and mass balance, instead. Further analysis on this quartic specification suggests that the effect of each individual elevation term is reasonably constant across the sample space, and therefore, adjusting for potential heteroscedasticity, OLS is the best linear unbiased estimator of this functional form. Also, mass balance profiles are computed and presented (Table 7) and their effects presented as well (Fig. 5). Throughout this study, absorbed solar radiation at a point was controlled for with the *norm*, regressor. Its effect is highly variable from year to year, likely as a result of variable weather from year to year. Regardless, it is a primitive regressor.

The data was likely collected by manual stake measurement methods. This type of measurement was taken by stabbing a glacier with a metal rod, and then coming back one year later to observe how much more or less of the rod was showing, accordingly, this yields the rate of either accumulation or ablation at a point on the glacier. These are not perfectly accurate measures, since sometimes, glacial forces would twist or destroy the rods.

II. Theory

Mass balance profiles are models that describe mass balance measurements for a given glacier for a given year. Historically, it is found by OLS on about 4 to 7 mass balance measurements by explaining mass balance as a cubic function of elevation above sea level.

I reason that mass balance at a point, b_i, can be described as the sum of ice and snow that accumulates and the sum of ice and snow that melts away; in other words:

$$b_i = accumulation + melt$$

Conceivably, accumulation at a point can be described by either precipitation or snow drifting into the given point. Melt, on the other hand, is described by snow drifting out of the given point, heat transferred to the point by conduction, heat transferred to the point by convection, and heat transferred to the point by radiation. That means that:

$$accumulation = precipitation + drift_{in}$$

and

$$melt = conduction + convection + radiation + drift_{out} + firnification$$

Here, mean annual precipitation is some function of climate, which in turn is some function of temperature, prec(T). Both rate of conduction, cd(T) and rate of convection, cv(T) are exponential cooling functions of temperature according to Newton's Law of Cooling, while firnification, the process by which snow condenses into ice, occurs at a rate as a function of temperature F(T). Now, radiation is a function of how much sunlight a given point is exposed to. This annual amount of sunlight at a point can be expressed by the amount the amount of absorbed incident radiation directed at a point, A, the amount of reflected solar radiation directed at a point, R, and the amount of shading that shields the point, S. Therefore,

$$accumulation = prec(T_i) + drift_{in,i}$$

and

$$melt = cd(T_i) + cv(T_i) + A_i + R_i + S_i + drift_{out,i} + F(T_i)$$

To describe absorbed incident radiation, I first consider a notebook placed in the sun. If I angle the notebook towards the sun, the sun shines on it. If I angle the notebook away from the sun, it lies in its own shadow, so the sun does not shine on it. The same should hold for a given point on a glacier. If a spot on a glacier has a normal vector parallel to the vector of incident solar radiation, than that spot should absorb sunlight better than an equivalent point that has a normal vector that is perpendicular to the vector of incident solar radiation, or a spot that lies in its own shadow. So, incident absorbed radiation, A, can be described as the dot product of the point's anti-normal vector, $-\vec{N}$, and the incident radiation vector, $\vec{I}$, with some proportionality constant, a, likely related to the point's albedo, or reflectivity index.

$$A_i = -a_i\vec{N}_i \circ \vec{I}_i = -a_i N_i I_i cos\phi_i$$

Since the exact mean value of $\vec{I}$ is unknown (although likely a some function of climate and macroscopic latitude), angle between these two vectors, ϕ, is unknowable. Therefore I approximate it A, with a solar exposure estimate, A', that supposes proportionality to how south facing a point is, or the anti-cosine of angle of aspect, ψ, (where a due north aspect is measured as 0°) and the slope, which is equal to the tangent of the angle of gradient, θ. Note that A' is not a perfect substitute for A[1]. So for some arbitrary real constant, β_A,

$$A_i' = \beta_A \cos(\psi_i)\tan(\theta_i)$$

Many of the parameters described in this section are functions of temperature. And it is of importance to note that one can look from a sandy beach up to a snowy mountain. This is to say, the higher up a place is, the colder it is. Therefore, elevation, z, can act as an instrumental variable for temperature. Let's say that temperature can be described as some function of elevation, g(z). Then, mass balance can be described as,

$$b_i = accumulation + melt$$

$$b_i = prec(T_i) + drift_{in} + drift_{out} + cd(T_i) + cv(T_i) + A_i' + R_i + S_i + F(T_i)$$

$$b_i = (prec + cv + cd + F)(T_i) + drift_{net} + A_i' + R_i + S_i$$

$$b_i = (prec + cv + cd + F) \circ g(z_i) + drift_{net} + A_i' + R_i + S_i$$

Standard procedure approximates $(prec + cd + cd + F) \circ g$ to a cubic polynomial. In this study I will approximate it to a polynomial, f, of variable degree to determine whether or not a cubic specification actually best fits this type of model. Expanding A' yields,

$$b_i = f(z_i) + A_i' + drift_{net} + R_i + S_i$$

$$b_i = f(z_i) + \beta_A \cos(\psi_i)\tan(\theta_i) + drift_{net,i} + R_i + S_i$$

Since the dataset does not include extractions from a snow drift model, this study cannot estimate the effect of the net drift at a point, nor can it estimate the effect of reflected radiation from surrounding areas for want of local albedo data. Furthermore, it cannot estimate whether or not each individual point is in the shade since the dataset lacks hillshade estimates. Therefore, these factors are lumped into the error term, u_i,

---

[1] A linear effect specification is used because in the preliminary study, other specifications such as a quadratic effect empirically proved drastically insignificant. Also, there are other theoretical issues with this primitive regressor addressed further in *Results.*

$$b_i = f(z_i) + \beta_A \cos(\psi_i)\tan(\theta_i) + u_i$$

But it is of significance to note that u is likely spatially correlated. The amount of reflected radiation incident on a point is a function of the geometry and reflectivity of surrounding area, so if surrounding ice is extra mirror-like, R will be larger. Similarly, S is likely spatially correlated, since if terrain stands in between a given point and the sun, S will have a larger effect. That is to say, for this dataset, if a mountain peak is immediately to the south of a given point, then that point will tend to be in the shade, and therefore the sun will tend to shine on it less. Therefore, it is reasonable to suppose that the inclusion of a spatial error correction model will improve the mass balance model.

$$b_i = f(z_i) + \beta_A \cos(\psi_i)\tan(\theta_i) + \lambda \boldsymbol{W} \xi_i + e_i$$

Where λ is a spatial autoregressive parameter, $\boldsymbol{W}$ is a spatial weighting matrix, ξ is the spatially autocorrelated error, and e is randomly distributed error.

III. Data

This dataset consists of a pooled set of 276 manual mass balance stake measurements of the Gulkana Glacier in central Alaska from 1966-1968. They original data was likely gathered by L.R. Mayo with the Water Resources Division fo the Fairbanks, Alaska branch of the USGS in the 1960s. I have no reason to believe that this data was ever analyzed. It was recovered in the Fairbanks library in December 2012. I personally digitized and estimated most of these values from printed maps with the use of ArcGIS. I believe that this is by far the largest data set of its kind, since most manual mass balance datasets include maybe five or six stake measurements per year, and therefore it likely has the ability to estimate effects and relationships that previously could not have been estimated due to lack of the sample size. The summary of the potentially relevant portions of the data set is as follows:

Variable	Obs	Mean	Std. Dev.	Min	Max
Balance	276	.2637681	1.465436	-5.4	3.6
Easting	276	579218.4	1749.053	575766.6	582618.8
Northing	276	7018105	1239.153	7014248	7019852
Z_interp	276	1862.354	206.2209	1203.785	2270.573
U_interp	276	22.6511	222.8137	2.559629	2625.87
aspect_64	276	196.1014	76.40326	0	358.32
slope_64	276	11.11868	5.191391	0	29.9006
Dist_Main	276	1226.101	924.2077	3.957915	3713.342
Dist_West	276	1645.23	846.9528	4.798514	3421.249
year	276	1966.993	.8484971	1966	1968

I describe the above variables as follows:

Variable	Units	Description
Balance	m	distance a glacier vertically ascended or receded in a given year
Easting	m	meters east of WGS 1984 UTM Zone 6N
Northing	m	meters north of WGS 1984 UTM Zone 6N
z	m	estimated glacial elevation for a given year. Interpolated from a Digital Elevation Model (DEM) of 1964 I made and a 1993 DEM computed from satellite data using a macroscopic time series model of the Gulkana Glacier computed by Louis Sass of the Anchorage USGS.
U	m	uncertainty in glacial elevation for a given year. Estimated from uncertainties in the 1964 DEM and the 1993 DEM.
aspect_64	degrees	direction of gradient with 0º as due north. Based on 1964 DEM.
slope_64	degrees	slope of gradient with 0º as perfectly flat. Based on 1964 DEM.
year	year	year in the set 1966, 1967, 1968

In this dataset, two observations proved unviable via leverage versus residual plots and box plots. Two had no value for both the 1964 DEM and the 1993 DEM. Accordingly, they are omitted, causing the final dataset to have 274 observations.

I developed the 1964 via standard kriging with a spherical autocorrelation model that used over 64,000 data that I manually extracted from a 1964 topographical map discovered with the dataset. It was coregistered against a 2001

DEM calculated from satellite imaging according to the method described in Nuth & Kaab, 2011 such to minimize any potential bias in the 1964 DEM.

In order to make the results more conceptual, the elevation value is rescaled in terms of kilometers.

$$z\ (m) * \frac{1km}{1000m} = \ kz\ (km)$$

Also, the regressor, *norm*, is calculated to represent absorbed incident radiation as follows:

$$norm = \ -\cos\ (aspect\_64) * \tan\ (slope\_64)$$

IV. Results

Determining the Functional Form of Mass Balance Profiles

In this section of the study, I use OLS to compute seven mass balance models as a polynomial function of elevation with degree varying from one to seven. Based on the estimated results and postregressive diagnostics (Table 1), I determine that a quartic function of elevation best describes mass balance. I determine this with the full 274 observations of the final dataset while controlling for both solar exposure and for temporal variation. I control solar exposure with the derived exposure regressor, *norm,* while I control for temporal variation with a battery of dummy variables derived from *year*. The temporal control essentially reduces each of these models to three separate models, one for each year, 1966, 1967, and 1968. Also, since several test models exhibited signs of autocorrelation and heteroscedasticity, all of these models (Table 1) preemptively correct for this with White's robust standard errors.

Table 1: Functional Forms of Mass Balance Profiles

VARIABLES	1st Balance	2nd Balance	3rd Balance	4th Balance	5th Balance	6th Balance	7th Balance
1967	-2.485**	8.709	228.4***	-177.6	-177.6	-252.5	-2170
1968	-0.172	-0.0525	26.98	-45.44	-45.44	-14.54	2000
kz	6.083***	21.87***	95.71***	-1,505***	-1,505***	-2291	-1823
1967*kz	1.450***	-11.51*	-406.7***	504.7	504.7	688.2	6146
1968*kz	0.0619	-0.867	-48.66	164.3	164.3	99.42	-5247
kz^2		-4.508***	-46.65**	1,331***	1,331***	2179	1674
$1967*kz^2$		3.673*	238.0***	-516.2	-516.2	-683.2	-6636
$1968*kz^2$		0.447	28.13	-188.7	-188.7	-138.4	5298
kz^2			7.909**	-513.3***	-513.3***	-943.3	-687.5
$1967*kz^3$			-45.85***	226.8	226.8	293.7	3281
$1968*kz^3$			-5.262	87.8	87.8	70.7	-2471
kz^4				73.19***	73.19***	164.4	110.1
$1967* kz^4$				-36.33	-36.33	-46.28	-642.4
$1968* kz^4$				-14.41	-14.41	-12.26	460.1
kz^5					0	0	0
$1967* kz^5$					0	0	0
$1968* kz^5$					0	0	0
kz^6						-1.892	-0.766
$1967* kz^6$						0	0
$1968* kz^6$						0	0
kz^7							0
$1967* kz^7$							3.528
$1968*kz^7$							-2.253
							1.666
1966*norm	-0.111	-0.671	-0.478	-0.221	-0.221	-0.216	-0.219
	0.469	0.414	0.402	0.369	0.369	0.37	0.369
1967*norm	-0.213	-0.236	-0.389	-0.358	-0.358	-0.358	-0.354
	0.429	0.514	0.489	0.447	0.447	0.448	0.446
1968*norm	-1.254***	-1.203***	-1.195***	-1.131***	-1.131***	-1.133***	-1.146***
	0.295	0.303	0.288	0.264	0.264	0.264	0.264
Constant	-11.04***	-24.59***	-67.06***	622.7***	622.7***	903.9	736.6
	0.741	2.791	18.79	101.7	101.7	570.2	632.8
Observations	274	274	274	274	274	274	274
R-squared	0.850	0.869	0.883	0.904	0.904	0.904	0.906
AIC	457.9	427.1	401.7	354.3	354.3	356	356.2
BIC	490.5	470.5	455.9	419.3	419.3	424.7	432
Ramsey's F	35.63	22.72	11.18	4.33	4.33	4.54	4.78
Ramsey's p	0.0000	0.0000	0.0000	0.0053	0.0053	0.0041	0.0030

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The effect of elevation in these models appears to decay within each model as the order of the effect increases, while across models, as the degree of the model increases, the magnitude of each elevation effect increases. There, as a tendency, appears to be insignificant difference between the effects of elevation between the three years, except in lower degree models (i.e. 1st, 2nd, and 3rd), there is a statistically significant difference in these effects between 1966 and 1967. Also, the effect of solar exposure is insignificant for 1966 and 1967, but consistently statistically significant on the 1% level for 1968. This suggests that the effect of this regressor is subject to temporal variation.

The lower degree models have reasonable parameters that appear subject to consistently statistically significant elevation effects, but these models also have very high F-statistics from Ramsey RESET tests, suggesting that there exists a model with a higher degree polynomial of the regressors that would better describe the data.

The Ramsey RESET F-statistic of the linear model (35.63) very strongly suggests a higher degree polynomial. This statistic decays in the quadratic, cubic, and quartic model and then begins to stagnate in the 4's and actually begins to increase from the quartic towards the septic model. So although the test consistently suggests a higher degree polynomial, it suggests this with the least vigor in the quartic and quintic models. As a quick caveat, these F-statistics are not all distributed the same since their numerator degrees of freedom vary with the number of regressors, which is variable across models, therefore they are not perfectly comparable. Large differences in their value, though, do reflect significant differences in test results.

The higher degree models, by contrast with the lower degree models, have unreasonable parameters, since the larger order effects tend to be economically insignificant (mechanically zero) and statistically insignificant.

The Ramsey RESET tests in conjunction with these parametric observations draws attention to either the quartic model or the quintic model. The quintic model is calculated to be equivalent to the quartic model since all 5th order effects are found to be mechanically zero, suggesting that a quartic model is the best mass balance model. This is supported by both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are both lower at the quartic specification then the linear, quadratic, and cubic specifications suggesting that the data is, in fact, best explained by this model. Admittedly, the information criteria are improved in the sextic and septic specifications, but they lack any statistically significant effects of elevation. The quartic specification has a reasonably higher R^2 value than the cubic model does. Moreover, all four elevation effects from 1966 are statistically significant on the 1% level and are all not different from the elevation effects of 1967 to 1968 to a statistically significant degree.

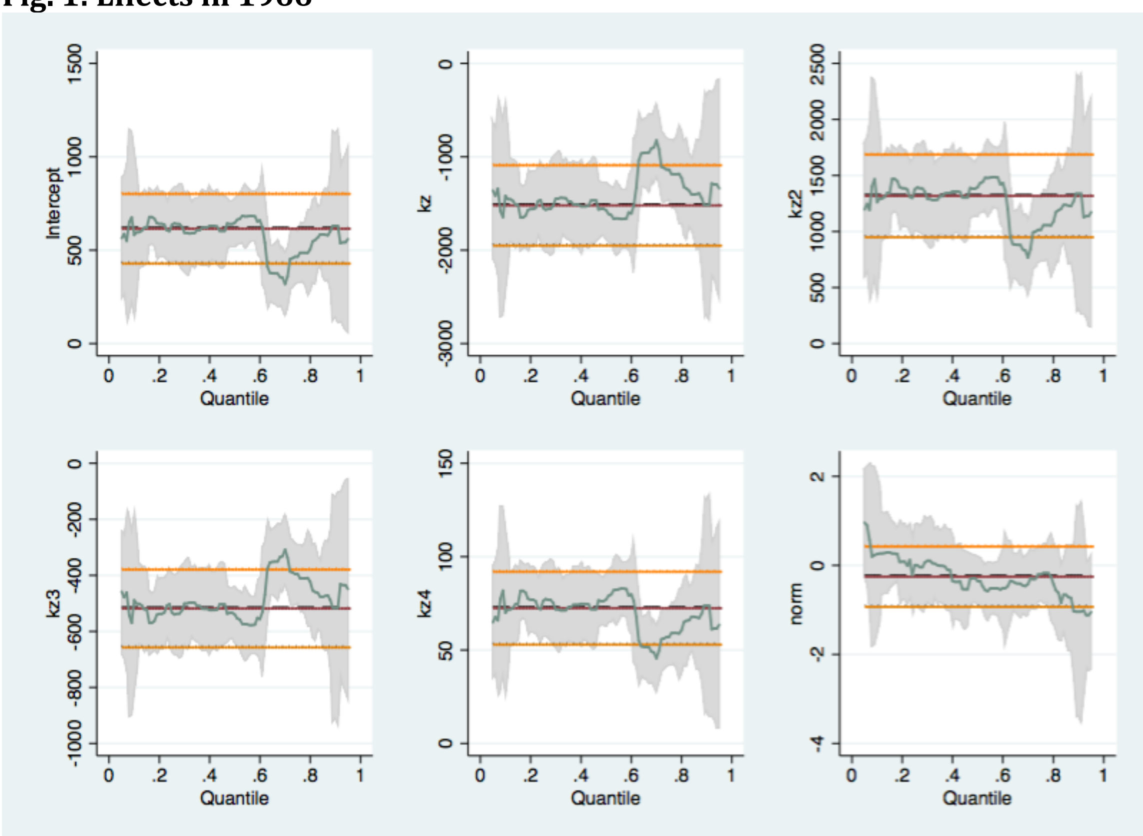
The effect of solar exposure is consistent and statistically significant on the 1% level for the year 1968 across models, but consistently insignificant for both 1966 and 1967. This suggests that the effect of sunlight varies annually, having an effect in 1968 but not in 1966 or 1967. This could perhaps be due to variable amounts of sunlight or maybe nearby volcanic activity.

Estimating Variations in the Effects on Mass Balance

In this section of the study, I estimate the effects of the regressors of the above quartic model over the sample space via quantile regression. Each year's model is calculated individually; that is to say, the data are not pulled together into a single dataset in this segment of the study.

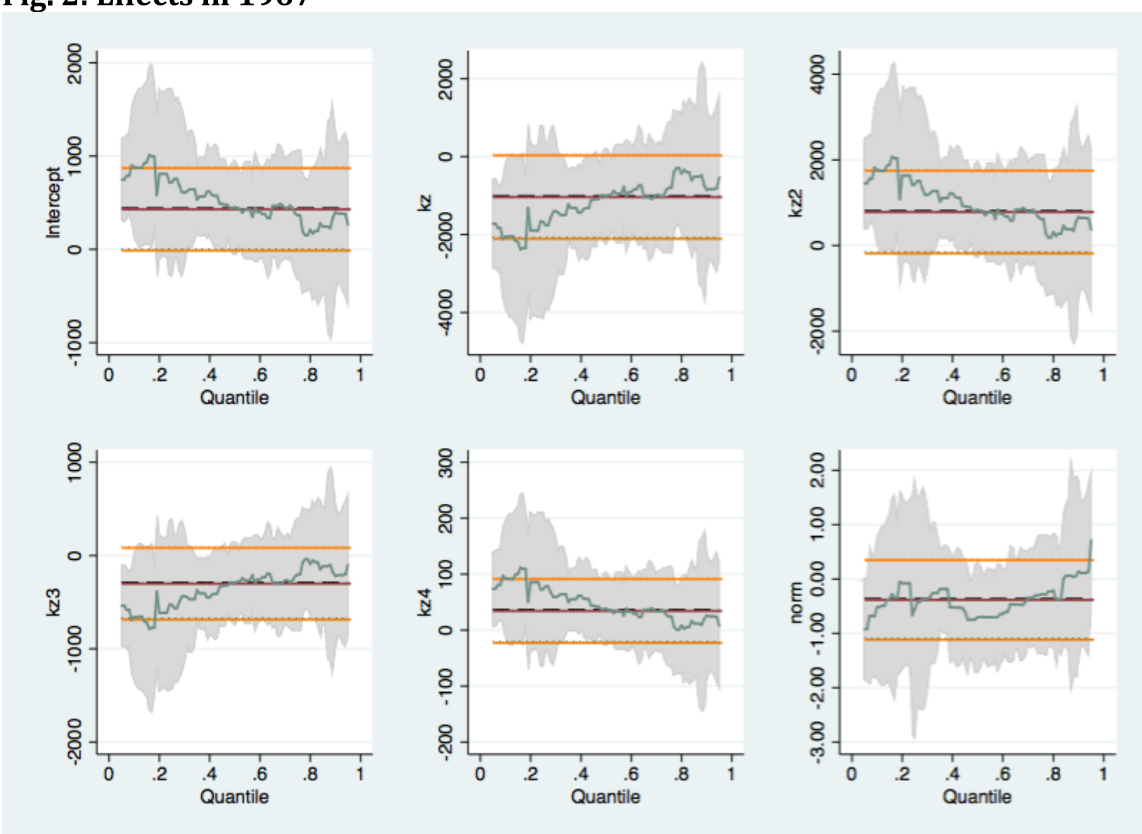
The intercept and effects of elevation in 1966 (green line) closely resemble the effects estimated by OLS (red line) in the first 60% of observations, which is roughly equal to all readings less than 0.78m. Also in this region, the 95% confidence interval of quantile regression (grey region) tends to fall near or around the 95% confidence intervals predicted by OLS (orange lines). After this region, though, in the 60% to 80% quantiles (mass balance from roughly 0.78m to 1.2m), there is consistently either a dip or spike in the intercept or effect of elevation that differ significantly from those predicted by OLS. After the 80% quantile, though, these effects of 1966 return to their predicted OLS values. These effects change drastically and immediately in all elevation effects and the interaction effect; this could perhaps be because of an underlying piecewise functional form where one trend holds in the first 60% of this data but a different one holds in the last 40%. This trend is likely specific to this year in particular, since it is not observed in either other year. Something to note, however, is how all five of these effects are either in perfect or in near-perfect compliment of one another. The shape of the distributions appear to be all equivalent across the sample space modulo sign and scale.

Fig. 1: Effects in 1966



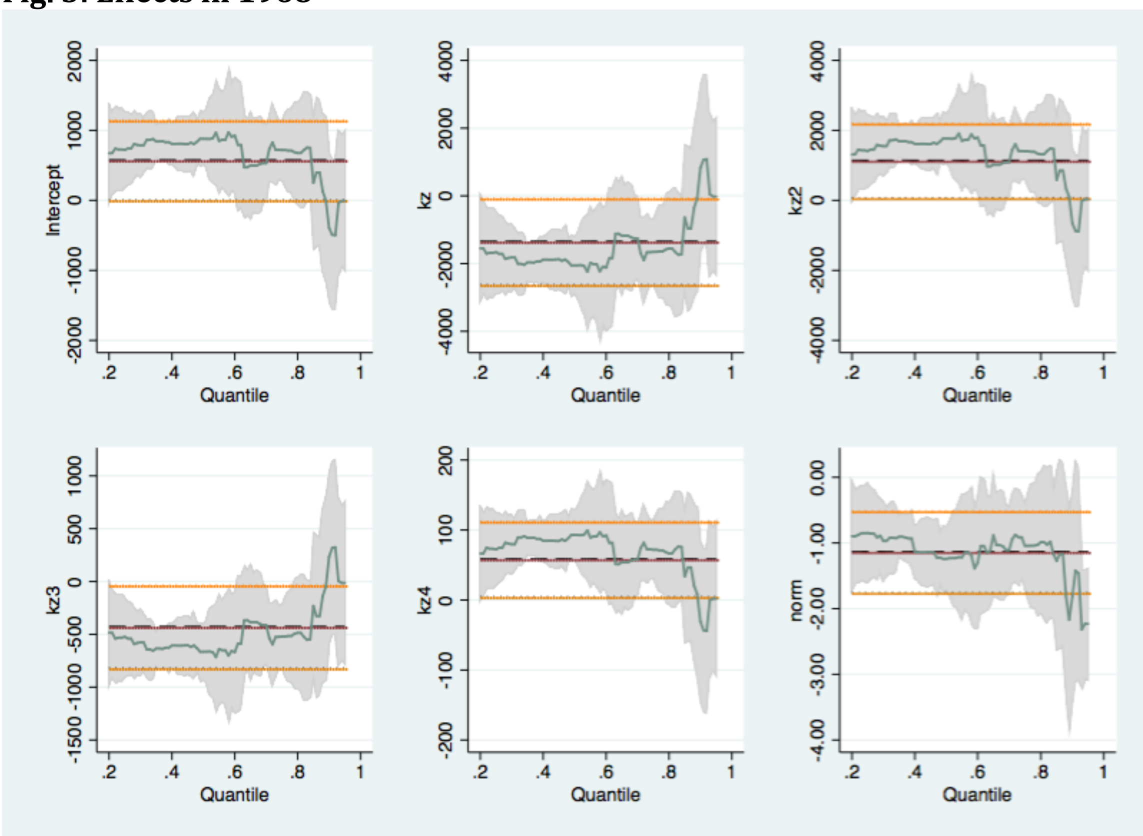
The intercept and effects of elevation in 1967 (green lines) appear to insignificantly depart from the intercept and effects of elevation predicted by OLS (red line). There does appear to be a loss of efficiency that results from quantile regression in the upper and lower regions of the distribution. This is because the width of the 95% confidence interval of quantile regression (grey region) is larger than that of the OLS (yellow lines). There does appear to be either an upward or downward linear trend in the intercept and effects of elevation over the sample space, but the quantile estimator with this dataset lacks the power to confirm its significance. This observable but insignificant upward/downward trend is repeated in neither 1966 nor 1968, therefore it is year-specific if it actually exists and does not describe any generalizable aspects of mass balance. Another observation worth note is that in 1967, as in 1966, the intercepts and effects of elevation are perfect compliments, the shape of the relations are the same. Modulo sign and scale, these parameters mirror each other; this suggests that these effects are interrelated and can be described by perhaps one parameter.

Fig. 2: Effects in 1967



The intercept and effects of elevation in 1968 from quantile regression (green) do not significantly depart from the values predicted by OLS (red), although a spike in test efficiency near the 40% quartile (mass balance of roughly 0.47m) almost significantly departs from OLS. Since this anomaly is not repeated in the other years, 1966 and 1967, I do not have significant reason to believe that it is more than just an anomaly. I truncated this quantile regression such to omit results from 1968 with quantile less than 20%. This was due to an explosion of standard errors in this region that dwarfed the other regions. The expected effects in this lower region did not appear to depart significantly from those expected by OLS, simply the precision here increased dramatically. In the upper end of the distribution, there does appear to be a large shift in the effects of quantile regression with respect to the effects of OLS, although this departure is not statistically significant. What is of note, though, is how the intercept term and the effects of elevation are, yet again, perfect compliments of one another; modulo sign and scale, these parameters follow the essentially the same shape. This could perhaps suggest the existence of a single estimable effect that would explain all five of these effects. As a further note, as a tendency, there appears to be a loss of efficiency in quantile regression at the high and low extremes of an effect distribution.

Fig. 3: Effects in 1968

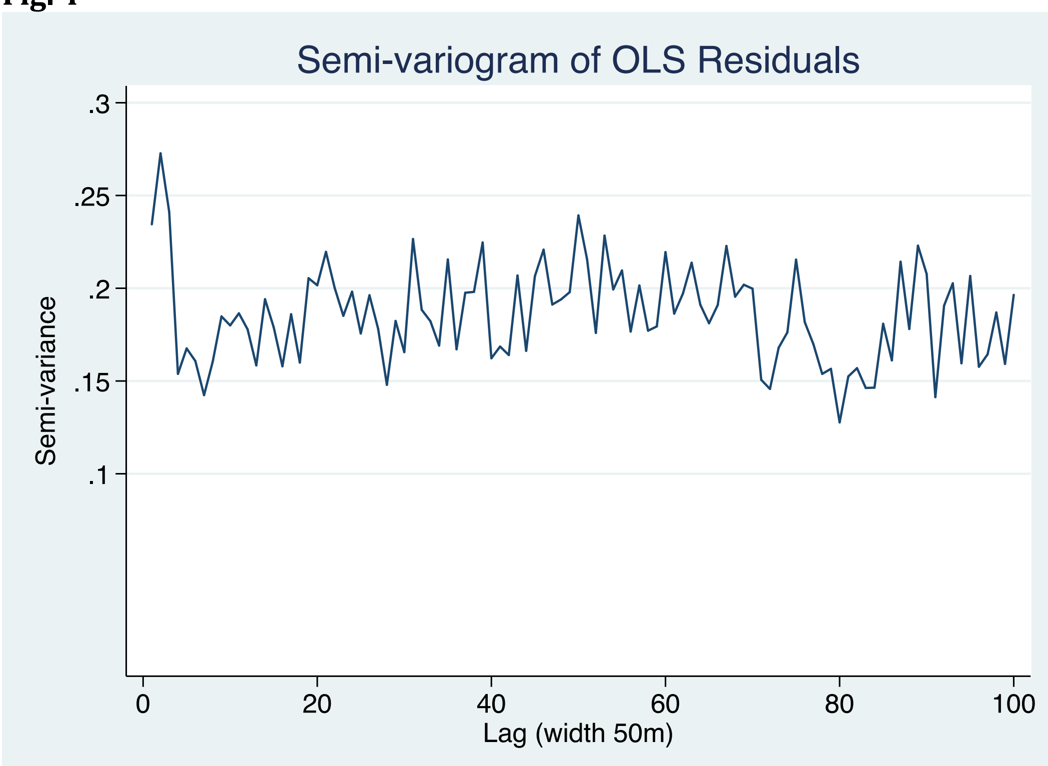


The effect of solar exposure predicted by quantile regression insignificantly departs from the predicted value of OLS for all 1966, 1967, and 1968. In 1966, there appears to be a slight downward tendency in expected value. In 1967, there appears to be an upward tendency with a loss of efficiency near the 20% and 90% quartiles. In 1968, there is an abrupt change in effect at the upper end of the distribution; I would suppose that in 1968, the quantile regression became subject to bias at the uppermost quartiles, and that this is why solar exposure, as the other parameters, changed abruptly but insignificantly from those of OLS. All in all, OLS appears to be the more efficient estimator for this parameter, solar exposure, and is therefore recommended. The effect of solar exposure does not compliment the other effects as the other effects do eachother, and therefore, solar exposure appears to have a separate effect from the intercept term and the effects of elevation.

Estimating the Mass Balance Profiles for the Gulkana Glacier from 1966 to 1968

In this section of the study, I apply a quartic function of elevation to the formula derived in section three in order to build the optimal mass balance profiles for the Gulkana Glacier in 1966, 1967, and 1968. As discussed in theory, I would expect localized spatial autocorrelation to be present due to omitted variable bias from lack of a hillshade regressor and lack of a snowdrift regressor. To visually look for spatially autocorrelated errors I calculate the semi-variogram (Fig. 4) of the residuals from OLS specified the quartic model (Table 1, 4th). Since semi-variance is effectively stationary over the lag space, there is no visually detectable spatial autocorrelation of error.

Fig. 4



To confirm this visual finding, I calculate an inverse-distance weighting matrix with a maximum spatial lag of 4 kilometers and use it to test the errors used in the aforementioned semi-variogram for autocorrelation. Using a one-tailed Wald test on Geary's *c* statistic, I find that there is insignificant localized spatial error autocorrelation ($p \approx 0.39$). Using a one-tailed Wald test on Moran's *I* statistic, however, I find that there is significant global spatial error autocorrelation ($p \approx 0.03$).

Table 2: Autocorrelation Test Results

Moran's I

Variables	I	E(I)	sd(I)	z	p-value*
ehat	**0.023**	**-0.004**	**0.013**	**1.967**	**0.025**

Geary's c

Variables	c	E(c)	sd(c)	z	p-value*
ehat	**1.014**	**1.000**	**0.050**	**0.287**	**0.387**

*1-tail test

So supposing that there is spatially autocorrelated error, I calculated the quartic specification (Table 2, 4th) using a maximum likelihood estimator with a spatial error correction model defined by an inverse-distance weighting matrix. The spatial autoregressive parameter of this model is $\lambda \approx -0.00116$. This value is statistically insignificant by all a Wald test ($p \approx 0.24$), a likelihood ratio test ($p \approx 0.24$), and a Lagrange multiplier test ($p \approx 0.12$).

Table 3: Tests on λ

Wald test of lambda=0:	chi2(1) =	**1.405 (0.236)**
Likelihood ratio test of lambda=0:	chi2(1) =	**1.401 (0.237)**
Lagrange multiplier test of lambda=0:	chi2(1) =	**2.394 (0.122)**

So, since I cannot visually observe any spatial autocorrelation, localized spatial autocorrelation of error is insignificant, and the inclusion of a spatial error correction model yields insignificant changes to error, it is safe to say that there is no significant spatially autocorrelated error in this dataset.

In order to confirm the validity of OLS as the best estimator for this data, the error term of the quartic OLS specification of the model (Table 2, 4th) must be normally distributed and homoscedastic. A skewness and kurtosis test (Table 4) for normality suggests that the error of the OLS model is not normally distributed ($p<0.00$). Moreover, a Breusch-Pagan test (Table 4) suggests the presence of heteroscedasticity ($p<0.00$). This is corrected for with White's robust standard errors (Table 5).

Table 4: Test for Normality and Heteroscedasticity

```
. sktest ehat

                    Skewness/Kurtosis tests for Normality
                                                          ------- joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
        ehat |        274     0.0026        0.0000        35.22       0.0000
. estat hettest err, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: err

         chi2(1)      =     8.46
         Prob > chi2  =   0.0036
```

I apply OLS to the mass balance profile described by a quartic polynomial function of elevation, while correcting for solar exposure and applying a battery of annual dummy variables to correct for temporal differences in the data (Table 1, 4th). The results of this model (Table 5) yield a statistically significant constant term of 1966, linear effect of elevation in 1966, quadratic effect of elevation in 1966, cubic effect of elevation in 1966, and quartic effect of elevation in 1966. Also, there is no statistically significant difference between any of these five effects with those of either 1967 or 1968. Again, the effect of solar radiation is significant for 1968 but not for either 1966 or 1967. Overall, the model appears to describe the data rather well ($R^2 \approx 0.90$).

Table 5:

Variable	Effect on Balance
1967	-177.6
	(223.5)
1968	-45.44
	(272.7)
kz	-1,505***
	(247.7)
1967*kz	504.7
	(536.9)
1968*kz	164.3
	(616.3)
kz^2	1,331***
	(206.9)
1967* kz^2	-516.2
	(480.3)
1968* kz^2	-188.7
	(517.9)
kz^3	-513.3***
	(76.19)
1967* kz^3	226.8
	(189.8)
1968* kz^3	87.80
	(192.0)
kz^4	73.19***
	(10.44)
1967* kz^4	-36.33
	(27.94)
1968* kz^4	-14.41
	(26.49)
1966*norm	-0.221
	(0.304)
1967*norm	-0.358
	(0.345)
1968*norm	-1.131***
	(0.276)
_cons	622.7***
	(110.3)
R^2	274
N	0.904

Standard errors in parentheses; *** $p<0.01$

Since there is no significant interannual difference effect in this model, I perform an F-test comparing this model to one where all temporal difference effects are zero (Table 6). This test suggests that although not a single one interannual difference effect is significant, eliminating all of them would result in a significantly weaker model (p=0.0005). Since I cannot eliminate all of difference effects and maintain the model's viability, and since I cannot eliminate a few of the difference effects without introducing bias into the model as a result of choosing one difference effect over another, I must keep all of the difference effects in the model, even though they are not individually significant.

Table 6: F-test on temporal difference effects

```
( 1)  1968.year#c.kz = 0
( 2)  1967.year#c.kz#c.kz = 0
( 3)  1968.year#c.kz#c.kz = 0
( 4)  1967.year#c.kz#c.kz#c.kz = 0
( 5)  1968.year#c.kz#c.kz#c.kz = 0
( 6)  1967.year#c.kz#c.kz#c.kz#c.kz = 0
( 7)  1968.year#c.kz#c.kz#c.kz#c.kz = 0
( 8)  1967.year#c.kz = 0

      F(  8,   256) =    3.63
           Prob > F =    0.0005
```

In this model (Table 5), the constant intercept term has no valuable interpretation, since the glacier would not have grown 623 m if it were sea level in 1966. The 1967 and 1968 terms describe the difference between the intercept term of 1966 and the intercept term for 1967 and 1968, respectively. These two values are insignificant, therefore the intercepts for 1967 and 1968 in this model are not significantly different from that of 1966. The effect of incident solar radiation here is statistically significant for both 1966 and 1967 but statistically significant in 1968 on the 1% level. Its 1968 effect is exactly significant in magnitude, though. A one standard deviation increase in *norm* (0.1469) yields a 0.16 m drop in mass balance, which is about a 0.12 standard deviation decrease in mass balance. Admittedly, moving from a vertically east facing point to an otherwise equivalent south facing point with a 45° gradient would suggest an expected 1.13 m decrease in mass balance, about a 0.81 standard deviation decrease in mass balance. Although this seems like a significant impact, it merely highlights the issue the replacement estimator for absorbed incident radiation has with extreme values.

The effect of elevation in this model is more complicated to interpret. Since mass balance is a higher degree polynomial of elevation, the effect of elevation varies over the sample space. That is to say, since,

$$b_i = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \beta_3 z_i^3 + \beta_4 z_i^4 + A_i' + u_i$$

it must be the case that,

$$\frac{\partial b_i}{\partial z_i} = \beta_1 + 2\beta_2 z_i + 3\beta_3 z_i^2 + 4\beta_4 z_i^3 \quad (1)$$

But also, the expected effect of elevation on mass balance can be easily computed for a given elevation. For example, the expected effect of elevation on mass balance for the above model (Table 5) computed at the mean elevation (1.87 km) is 4.45 m/km with a standard error of 0.38 m/km. This value is statistically significant, but also is reasonably significant in magnitude, since a one standard deviation increase in elevation (0.20 km) here would entail an expected 0.89 m increase in mass balance, which is a 0.63 standard deviation in mass balance. Computing separate OLS models for each year (Table 7), though, estimates the effect of elevation (with standard errors in parentheses) at the mean elevation to be 2.70 (±0.46) m/km for 1966, 8.25 (±0.71) m/km for 1967, and 4.05 (±0.79) m/km for 1968. The mass balance profiles here (Table 7) were calculated with White's robust standard errors because a Breusch-Pagan test detected heteroscedasticity in the error of the 1966 profile (p<0.00), the 1967 profile (p≈0.04), and the 1968 profile (p<0.00). Moreover, a joint skewness-kurtosis test suggests that the errors of these profiles are all not normally distributed (p<0.00); this should also be corrected by the use of White's robust standard errors. The elevation effects are not constant over the sample space, though.

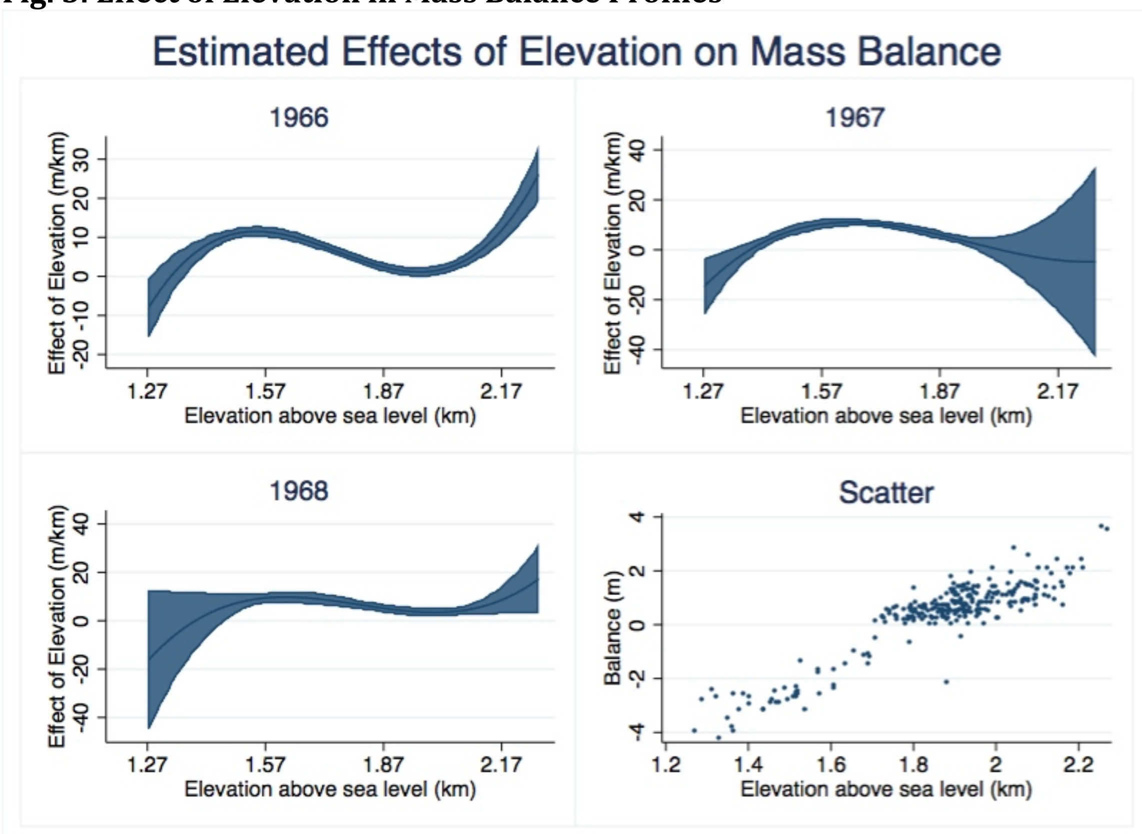
Table 7: Mass Balance Profiles for 1966-1968

	1966 Balance	1967 Balance	1968 Balance
kz	-1,505.348	-1,000.688	-1,341.066
	(6.09)**	(2.09)*	(2.38)*
kz^2	1,330.957	814.759	1,142.270
	(6.45)**	(1.87)	(2.41)*
kz^3	-513.255	-286.418	-425.452
	(6.76)**	(1.64)	(2.42)*
kz^4	73.188	36.861	58.777
	(7.03)**	(1.41)	(2.42)*
norm	-0.221	-0.358	-1.131
	(0.73)	(1.03)	(4.10)**
_cons	622.730	445.155	577.286
	(5.66)**	(2.27)*	(2.32)*
R^2	0.92	0.95	0.83
N	100	77	97

t-scores in parentheses * p<0.05; ** p<0.01

The effect of elevation on mass balance in each of the three profiles (Fig. 5) is represented by a cubic relationship according to Eq. 1 using the coefficients calculated in the original profiles (Table 8). The effect relationships of all three years (Fig. 5) appear relatively precise in the center of the elevation distribution, having very small 95% confidence intervals (blue region) in these regions of highest data density. The extrema of the elevation distribution tend to yield less precise elevation effect estimates, though, perhaps as a result of decreased data density. This anomaly is especially pronounced near the supremum of the 1967 elevation distribution, predicting an effect that lies in the 95% confidence interval of -40 m/km to 30m/km.

Fig. 5: Effect of Elevation in Mass Balance Profiles



V. Conclusion and Validity Assessment

This study suggests that a quartic polynomial of elevation best models mass balance profiles and that the distributions of the calculated coefficients from this polynomial are essentially equivalent to the coefficient distributions produced by ordinary least squares regression. This suggests that a quartic OLS regression of elevation produces the best mass balance profile. Furthermore, this study corrected for absorbed incident solar radiation, A', and found that its effect varies from year to year, which is to be expected if the total amount the sun shines varies from year to year. And when the effect of A' is statistically significant, it is not very significant in magnitude (beta coefficient = 0.12). And although theory suggests localized spatial autocorrelation of error, a static semi-variogram, a Wald test on Geary's c ($p \approx 0.39$), a Wald test ($p \approx 0.24$), a likelihood ratio test ($p \approx 0.24$), and a Lagrange Multiplier test ($p \approx 0.12$) on the maximum likelihood estimate for the spatial autoregressive parameter all suggest that there is no statistically significant localized spatial autocorrelation of error.

With respect to internal validity, the results presented in this study appear perfectly valid. There is perhaps a problem of multicolinearity between the different elevation terms in the pooled mass balance model and the individual mass balance profiles, but this will be inherent for any higher order polynomial function specification. This issue cannot be statistically adjusted for and can only be addressed by increasing the dataset. This only means that the regressed coefficients will be imprecise. The calculated effect of elevation at a given elevation appears to be reasonably unaffected by this inflation of variance, though. A skewness-kurtosis test for normality suggests that the errors of the pooled mass balance model and the individual mass balance profiles do not have normally distributed error. Similarly, a Breusch-Pagan test suggests heteroscedasticity in these models. I correct for this by application of White's robust standard errors to the models. In the pooled mass balance model, any potential for temporal autocorrelation of error is addressed by use of a battery of annual dummy variables used on every regressor. Omitted variable bias could only conceivably occur due to the omission of shading, reflected radiation, and snow drift variables, since they do not exist in this dataset. There is no reason to believe that these variables would cause a significant omitted variable bias, though, since they would cause a localized spatial autocorrelation of error, which was extensively tested for and not found. Measurement error is likely insignificant in this study. Mass balance measurements are either taken to the nearest centimeter or the nearest decimeter when in the ablation zone. Elevation estimate error was estimated to have a mean value of about 3.6m, which is rather accurate since elevation is on the scale of kilometers. So the effect of this minimal measurement error, if there is any, is a bias towards insignificant regressor effects. Sample selection bias is likely not a significant issue in this study. Two data from the initial dataset were omitted, and this decision was based on their invalidity of their extracted values from both the 1993 DEM and the 1964 DEM. Their omission is determined by the explanatory variables, not dependent variable, and therefore sample selection bias is not an issue. There could perhaps be an issue of simultaneity bias here, since in the long run mass balance, which is like the rate of change of elevation, determines elevation. This study uses elevation to determine mass balance, so a reverse causality could yield biased, inconsistent results. But there is no viable instrumental variable for elevation in this dataset, and therefore any issue of reverse causality in this study would be equivalently present in every mass balance study. An issue of functional form may be a problem in this study, though. Not with the form of the polynomial function of elevation, but instead with the form of the effect of absorbed incident radiation. This regressor is supposed to have a linear effect, which contradicts the prior that there exists some maximum effect for a certain optimal normal vector which diminishes as the vector departs from this optimal value. I tried to apply a quadratic effect in the preliminary study, but the quadratic term consistently proved incredibly insignificant, both statistically and in magnitude. Moreover, this simplified regressor impractically treats non-south facing terrain. It suggests that a virtually flat region that is slightly south facing should have opposite effect when compared to an otherwise equivalent region that is slightly north facing. This contradicts the prior that they should be treated similarly. Admittedly, only about 26% of the data have an angle of aspect north of either due east or due west, and since the effect itself is insignificant in magnitude when it is statistically significant, it is extremely unlikely that it biased any of the other results. On the contrary, I'd believe that it still aided the study to a small degree. In future studies, though, I recommend deriving a different estimator for absorbed incident radiation. Otherwise, there is no reason I can think of to question the internal validity of this study.

With respect to external validity, though, I cannot confirm that the findings in this study will carry over to glaciers that are not the Gulkana Glacier, nor can I confirm that the findings in this study will carry over to periods outside of 1966-1968. The profiles and coefficients derived here certainly are specific to the latitude and climate of this specific time and region. I have no reason to believe that distributions of estimated effects will not closely follow those predicted by OLS as shown in this study, and also I have no reason to believe that the viability of a quartic mass balance profile won't carry over to other glaciers in other time periods. And given the size of this data set compared to other similar datasets, I believe that the quartic specification could likely be applied to other profiles. Since I do not observe multiple glaciers in this study, and since I do not observe more than three years of data, I cannot provide discrete confirmation of the viability of this functional form.

I selected the quartic elevation specification because proved to increase R^2, decrease information criteria, minimize the results of a Ramsey RESET test, and provide statistically significant results. This section of the study did not account for spatial error, but as outlined above, spatial error proved insignificant in this study. As a caveat to the quartic model, it can potentially produce a mass balance profile with an elevation effect that decreases with elevation over a region, which contradicts the prior assumption that the mass balance monotonically increases with elevation. Admittedly, this issue is also present with the cubic specification.

This study also calculates mass balance profiles for the Gulkana Glacier from 1966 to 1968 using OLS with White's robust standard errors. The results of this calculation can be found in Table 7. The estimated effect of elevation (with standard errors in parentheses) at the annual mean elevation is 2.70 (±0.46) m/km for 1966, 8.25 (±0.71) m/km for 1967, and 4.05 (±0.79) m/km for 1968. This means that a one standard deviation increase in elevation (≈0.2km) will yield an expected 0.54m increase in mass balance in 1966, 1.64m increase in 1967, and 0.81m increase in 1968, which is a 0.38, 1.18, and 0.58 standard deviation increase in mass balance for 1966, 1967, and 1968, respectively. These values are indicative of the effect of elevation at the mean elevations (or mass balance ≈0.5m) and are not indicative of the overall effect of elevation on mass balance, which varies over the sample space.

An interesting observation of this study is how the effect of every elevation term in a tested polynomial model (Table 1) decays as the order of the term increases. And also, the effects of the interaction term and all four elevation effects of the quartic specification (Fig. 1; Fig. 2; Fig. 3) are consistently perfect compliments of each other across the sample space; this could perhaps suggest that they could be described as some single variable. These findings in both suggest the possibility that the polynomial terms here can be expressed as a convergent infinite sum. In *Theory*, I suggested that the polynomial function of elevation serves as an empirical proxy for several complex effects such as heat loss due to convection and conduction. Newton's Law of Cooling demands that an exponential term be used to describe cooling. An exponential term can also be expressed as a convergent infinite sum. Perhaps, something that looks like an exponential term could yield the optimal mass balance profile.