ECONOMETRIC EVALUATION OF LARGE WEATHER EVENTS DUE TO CLIMATE CHANGE: FLOODS IN ATLANTIC CANADA

by

Stanislav Hetalo

Bachelor of System Analysis (Taras Shevchenko National University of Kyiv, Ukraine, 2016)

A Report Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Arts

in the Graduate Academic Unit of Economics

Supervisor: Yuri Yevdokimov, Ph.D., Economics

Examining Board: Constantine E. Passaris, Ph.D., Economics, Chair
Phil Leonard, Ph.D., Economics
Mehmet Dalkir, Ph.D., Economics

This report is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

May, 2018

ABSTRACT

According to the 5th Annual Report of the Intergovernmental Panel on Climate Change (IPCC), climate change will increase the frequency of large weather events such as floods, storm surges, cyclones, hurricanes, high speed winds, thunderstorms, snowstorms, blizzards, extreme temperatures and others. All these events lead to a significant economic damage to property, infrastructure and human health. Historically Atlantic Canada has been vulnerable to flooding. Destructive consequences of the flooding have been seen in the past and are expected to occur in the future specifically as a result of ongoing climate change. The ultimate goal of this study is to establish a relationship between socio-economic, climate change as well as direct flood factors and economic loss from floods in Atlantic Canada. As the first step in reaching this goal, the present study evaluates probability of floods in Atlantic Canada due to hydrological as well as climatological factors first and then tests the hypothesis of an increasing frequency of floods in the future due to climate change. Comprehensive statistical analysis performed in this study is based on the data collected from Canadian Disaster Database, Database of Environmental Departments and Local Governments of Maritime Provinces of Canada and Statistics Canada.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

According to the 5th Annual Report of the Intergovernmental Panel on Climate Change (IPCC), as the world has warmed, that warming has triggered many other changes to the Earth's climate. Changes in extreme weather and climate events such as floods, storm surges, cyclones, hurricanes, high speed winds, thunderstorms, snowstorms, blizzards, extreme temperatures and others are the primary way that most people experience climate change. Human-induced climate change has already increased the number and strength of some of these extreme events. Over the last 50 years, we have seen increases in prolonged periods of excessively high temperatures, heavy rainfalls and snowstorms, and in some regions, severe floods and droughts.

Historically Atlantic Canada has been vulnerable to flooding. Destructive consequences of the flooding have been seen in the past and are expected to occur in the future especially as a result of ongoing climate change. The Canadian Disaster Database recorded more than 100 meteorological and hydrological disasters occurred in Maritime Provinces during period of 1900-2014. Floods appear to be the most frequent among all extreme weather events as presented in the following table.

Table 1: Extreme weather events in Canada's Maritime Provinces

Type of event	**Quantity of events in different years**							
	1990-2014		**1965-1989**		**1940-1964**		**1900-1939**	
	total	**annual**	**total**	**annual**	**total**	**annual**	**total**	**annual**
Flood	22	0.88	17	0.68	9	0.36	10	0.25
Winter storm	15	0.6	2	0.08	2	0.08	1	0.025
Hurricane, tropical storm	11	0.44	5	0.20	6	0.24	3	0.075

Severe thunderstorm	8	0.32	7	0.28	1	0.04	2	0.05
Storm surge	4	0.16	3	0.12	0	0	0	0
TOTAL	60	2.4	34	1.36	18	0.72	16	0.4

Source: Data Adapted from Burina 2017

The author of the table given above – Burina (2017) – identifies extreme weather as an event meets at least one of the following criteria (based on definition from Canadian Disaster Database): ten or more people killed; one hundred or more people affected/injured/evacuated or homeless; an appeal for national/international assistance; event of historical significance; significant damage/disruption such that community could not recover without assistance. More importantly, as seen from the above table, frequency of floods has been increasing over reported period.

Huber and Gulledge (2011) note that in general climate change means not only average changes in the so-called climate related variable such as, for example, annual temperature or precipitations but also shifts in the frequency and severity of extreme weather events. Moreover, they claim that these shifts produce the highest economic and social damage among all climate change impacts. With respect to floods, the above mentioned authors claim that "… recent examples of flooding and extreme rainfall should provide lessons on where flood control and emergency systems are most needed and how much the investments in preparation are worth. Additionally, extreme events represent data points that can improve trends and estimates of future risk…" Thus, the goal of our study is multi-dimensional. First, we test a hypothesis about climate change and potential breakpoint in flood attributes in Atlantic Canada. Then we investigate how climate change influences on probability and frequency of flooding in the given region. Finally, we would like to consider the reliability of our outcome.

The rest of the study is organized as follows. The next chapter provides a brief overview of literature review on the given topic. Methodology description and general modelling approach is described in chapter 3. Then we describe our data and empirical framework along with key findings for econometric evaluation of flooding events in chapter 4. Chapter 5 concludes and provides insights in the potential future researches in the given area.

Chapter 2. Literature review

Based on research of the field, I could highlight two main conceptual streams in the literature with respect to given field. First one is based upon consideration of the flood from hydrological point of view. That's why I would like to classify this stream as "**hydrological**". Why it is important for our study? The main idea is to define variables that affect probability of floods. They might be divided in several groups: hydro-climatic, economic, physical, etc. Our idea is to define relevant variables in our case study. There are many specific variables might be used to build a model for flood forecasting, such as soil moisture capacity, rainfall intensity and rainfall depths for different periods of time etc. The most prominent model that is used to operate within given framework is called "rainfall-runoff model" or simply runoff model. It is a mathematical non-linear model that describes relations between drainage basin, catchment area and watershed. To be more precise, it produces runoff hydrograph as an output in response to inputs created by rainfall event. Let's take a look at this model in some specific literature scope.

Hlavcova et al. (2005) consider conceptual rainfall-runoff model with real-time estimates of the soil moisture conditions in a catchment. They state that such models usually are run by forecasting agencies and, as a result, data, methodology and results are publicly available. Their method is scenario-based. It captures all necessary results from an analysis of the return periods. Their rainfall-runoff model is non-linear and produces great results with respect to flood forecasting. These results were proven to be concise in Hron River Basin case study.

Vaze et al. (2012) suggest that complex methods that are being used for the rainfall-runoff model tend to be data-demanding while more simpler methods might generate heavy bias in the output. They also state there is exist need in "development more sophisticated approaches for parameter sets derived from multiple gauged catchments” and development of more progressive methods for rainfall-runoff models in the region scale.

Zhijia et al. (2008) perform rainfall-runoff simulation and then they make flood forecasting for Huaihe Basin. Their model was created to forecast the discharge hydrographs and channel routing with rainfall-runoff model itself. They made a real-time correction and considered possible influences of major flood gates as well as main reservoirs. After model calibration, it was applied to the flood of 2005 (while model was created based on the data from 2001 to 2004). The forecast of the model is proved to be accurate and consistent with observed results. Besides the reliable results, authors claim that “rainfall-runoff and flood forecasting based on hydrological models is complicated”. Hence, they made a few simplifications and did not include operation of flood diversion and retarding areas as well as some of the gates in the river networks. Inclusion of such data might have an integrated effect on the result while it makes given model even more complex and tangled.

There are a lot of other studies that consider runoff model: Calver and Lamb (1999), Cameron et al. (2000), Vaze et al. (2011a), Madsen (2000), etc. Most of them highlight high precision of the results and specify that such sort of modelling is resource demanding.

Hence, I would like to conclude that forecasting of floods from hydrological point of view in general and rainfall-runoff model in particular has an advantage in terms of great precision and reliability of the outcome. In this model floods are considered in their nutshell and results are based upon deliberate examination of the relevant variable and their interaction. Despite these huge pros, operation of this model requires deep understanding of the process itself and variables that might be relevant to a flood in each geographical area. This argument and complexity of the model create some significant barriers for researchers who do not have substantial hydrological knowledge. The main idea we took from this stream is a relevant set of potential factors that might relate to flooding, such as level of precipitations or river discharge.

My area of expertise lies outside hydrology. In this work, I am looking for much more transparent model that might be performed and understood by people without hydrological background. To create the model that satisfies given requirements, I have considered another big chunk of literature. A lot of authors use only commonly known variables for flood estimation (annual rain or snow precipitation, water discharge etc.) but experiment with different functional forms and/or statistical approaches. Various types of data, such as cross-sectional, time-series, or panel data could be used. Straightforward frameworks in this case are usually based on linear models, while more complex ones advocate for non-linearity in specifications. Probabilistic outcomes are estimated with wide range of models, such as logit, probit, generalized extreme value functions, etc. There is also a diversity in parametric and non-parametric modelling. As a result, I would like to classify this stream of literature as "**statistical**". Here we are looking for the best possible specification and estimation algorithm for our case study.

Hasanah, Herlina, and Zaikarina (2013) consider a transfer function model of water discharge and rainfall in Katulampa dam (Indonesia) to make a flood prediction system. Their model connects the output series, the input series, and white noise together. Despite the fact the model is comprehensive and requires a plenty of preparatory identifications and estimations, it provides significant results and shows that there is exists a link between floods, water discharge, and rainfall. In the end, authors of the article claim it would be more accurate if the more variables were included in the model. This question would be addressed in our work later.

Aich, Kone, Hattermann, and Paton (2016) perform time-series analysis of floods across Niger River Basin using hydro-climatic variables, such as precipitation level and annual maximum discharge (AMAX). Since Niger River has a huge basin, data was split across subregions for a higher precision of estimation. Non-stationary generalized extreme value functions and some non-linear methods were used to analyze data itself, change in the variance of annual flood peaks and damage caused by the flood. Their results show a strong relationship between floods, precipitation, and AMAX. Authors have also found strong relationship between increasing trend in AMAX and increasing flood magnitudes which indirectly increase flood risk.

Mohr, Kunz and Keuler (2015) develop and apply special logistic model to estimate the past and future hail potential in Germany. Despite the fact, hail attributes are different from flood determinants, the model itself is interesting for our purposes. The probabilistic outcome of the estimation of the model has a certain threshold to determine occurrence or absence of the given event in the considered year. They assert "The logistic model approach based on logistic regression improves the diagnostic of hail events…". However, like any model, their model has its own weaknesses, one of them is nature and availability of some data.

Mouri et all. (2013) describe a model for probability of assessment of floods in Japan. They introduce special flood risk index as a proxy for probability of floods in a given geographical area. To predict index value authors put into use the most essential climatic variables with respect to floods: precipitation and river discharge. To find the calculation of the return period for a given variables the data were assumed to follow

Gumbel distribution. They highlight that expected value of flood damage and economic losses might be estimated with higher precision using their index.

The literature on this topic is wide and more valuable ideas could be found in: Kundzewicz et al. (2013), Paeth et al. (2010), Casse et al. (2015), Sarr et al. (2013), etc. Their views are broad in nature. Overall, they have supplied us with a realm of ideas related to our case study.

So, based on the information from literature above, I would like to summarize that statistical stream of literature shows ways that allow people without hydrological background performing research in this field. This point and clear outcome of the framework for policy recommendation purposes are considered to be advantages of the statistical stream. In fact, smaller number of independent variables might lead to omitting some of them that might be relevant in a specific geographical area. This disadvantage might intensify level of biasness of the model furthermore.

As we can see from literature discussed the most common type of data used is time-series. Significant number of authors address non-linearity in their frameworks. Generalized linear models for binary dependent variables, such as logit are widely used in these cases. We would like to keep these points in mind during our modelling process.

Main purpose of this work is to create pellucid framework which does not require big number of people to operate or some very specific knowledge with respect to a given sphere. The model also should not be extensively comprehensive and results of it should be intelligible for people from various fields. In the end, this work is intended to become a part of a bigger general equilibrium (GE) framework which requires additional flexibility of outcome with respect to possible changes in independent variables set.

Hence, I would like to follow statistical stream of literature during my model specification with some points of hydrological to set up and estimate the model for a flood forecasting in Atlantic Canada.

Chapter 3. Methodology

The ultimate goal of this research is to show that climate change increases probability of floods in Atlantic Canada. Since the goal of this study is twofold, I divide it in two steps. In step one, we are going to find the best statistical model to explain probability of floods in Atlantic Canada as a function of socio-economic, climate and hydrological variables. In the end, we choose the most important variable. In step two, we analyze dynamics of the most significant climate variable to explain probability of flood - rainfall. Time-series analysis is used to testify for climate change and find evolutionary dynamics of the variable of choice from step one. In addition, we would like to test for endogenous break the rainfall variable. The two steps combined provide us with probability of future floods due to climate change as seen through changes in dynamics of rainfall. Burina (2017) statistically found the so-called damage function with respect to floods in Atlantic Canada. Coupled with future probabilities, the damage function can produce expected loss from floods. As already mentioned, all these outcomes of the study will give policy makes the upper bound for the investment in flood mitigation measures.

In the step one I consider 3 models that are used for data in binomial form: probit, logit, and complementary log-log transformation. I would like to use description for probit and logit models from the book "Econometric theory and methods" Russell Davidson and James G. MacKinnon (2004).

Consider framework of the binary response model in the following form:

$$y_t = X_t\beta + \mu_t \qquad (1)$$

In a model with binomial response, the value of the dependent variable $\mathbf{y_t}$ can take on only two possible values: zero or one. Denote P_t as the probability that y_t =

1conditional on some information set Ω_t. Then binary response model serves to model this conditional probability. Since for the dependent variable there are only two possible options: 0 and 1, it is clear that P_t is also expected value of y_t conditional on information set Ω_t:

$$P_t \equiv Pr\ (y_t=1 \mid \Omega_t) = E\ (y_t \mid \Omega_t) = X_t\beta$$

Obviously, any potential binary response model must satisfy that $0 \leq E\ (y_t \mid \Omega_t) \leq 1$. Even if this condition is held for all observations in a given sample, it is possible to find values of X_t for which our estimated probability value $X_t\hat{\beta}$ would contain numbers outside the 0-1 probabilistic interval. Theoretically, there are many possible ways how to ensure that given probability would be in the necessary 0-1 interval. In practice, there are not so much models are widely used. All of them ensure condition $0 \leq P_t \leq 1$ by specifying that:

$$P_t \equiv E\ (y_t \mid \Omega_t) = F\ (X_t\beta)$$

where $X_t\beta$ – index function, which maps vector of independent variables and vector of parameters into some scalar index, and $F(x)$ – is a transformation function, which should satisfy the following properties:

1) $F\ (-\infty) = 0;$
2) $F\ (+\infty) = 1;$
3) $f\ (x)\ =\ \frac{dF(x)}{dx} > 0.$

The first possible option is the cumulative standard normal distribution function:

$$\Phi\ (x)\ \equiv\ \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}X^2}\,dX$$

When transformation function $F\ (X_t\beta) = \Phi\ (X_t\beta)$, the model is called the probit model.

The logit model is quite similar to probit, but the transformation function is called logistic function and it has the following form:

$$\Lambda\,(x)\ \equiv\ \frac{1}{1\ + e^{-x}} = \frac{e^{x}}{1 +\ e^{x}}$$

Oosterbaan (1994) in their "Chapter 6: Frequency and regression analysis of hydrologic data" in Ritzema H. P. "Drainage Principle and Applications, Publication 16" describes the standard Gumbel distribution as

$$\Theta\ (x) \equiv e^{-e^{-x}}$$

This distribution also known as Generalized Extreme Value distribution Type - I and could also be considered as transformation function.

In fact, binary response models mentioned above are special cases of broader class of the so-called Generalized Linear Models (GLM). To be consistent with classification of the broader class, modern literature and software introduce concept of link function and apply it to the left-hand side of equation (1) instead of applying some transformations to the right-hand side. To explain theory behind it, I would like to follow German Rodriguez's (2017) lecture notes on GLM from Princeton University.

He states, "In fact, any transformations that maps probabilities into the real line could be used to produce a generalized linear model, as long as transformation is one-to-one, continuous and differentiable". So, let's consider *F(.)* – cumulative distribution function (c.d.f.) of some stochastic variable defined overall real line:

$$\pi_{\ i} =\ F\ (\eta_i)$$

where our random variable $-\infty < \eta_i < +\infty$. Then inverse transformation

$$\eta_i\ =\ F^{-1}\,(\pi_i)$$

for π_i within probabilistic interval 0-1 is called link function.

This concept became popular by introducing models for binary data connected with *latent variables*. Denote our dummy variable as Y_i. Assume that there is exist another unobservable stochastic and continuous variable Z_i which is defined on the entire real line. Consider our variable Y_i as an indicator which is equal to one if Z_i exceeds some given threshold θ. In such a case Z_i is called latent variable. Hence,

$$\pi_i = \Pr\{Y_i = 1\} = \Pr\{Z_i > \theta\}$$

We don't observe this variable, thus properties of latent variable Z_i are set by researcher. Popular practice in model identification of latent variable is to set threshold equal to zero and standardize Z_i to have standard deviation equal to some fixed number (for example, one).

Assume that outcome depends on a vector of covariates x. Let's write this dependence in the following way:

$$Z_i = x_i'\beta + \mu_i$$

where β is a vector of coefficients and μ_i is an error term that have some distribution with c.d.f. $F(\mu)$. According to this model:

$$\pi_i = \Pr\{Z_i > 0\} = Pr\{\mu_i > -\eta_i\} = 1 - F(-\eta_i)$$

where $\eta_i = x_i'\beta$ is the linear predictor. If the distribution is symmetric around zero, then $F(\mu) = 1 - F(-\mu)$. From here:

$$\pi_i = F(\eta_i)$$

This equation defines a GLM with binary response and the following link function:

$$\eta_i = F^{-1}(\pi_i)$$

If the distribution of the error term is not symmetric around zero:

$$\eta_i \; = \; -F^{-1} \, (1 - \; \pi_i)$$

There are 3 popular link function that are widely used in the literature and practice. Two of the them are based on the symmetric c.d.f. (logit and probit), while the third one is not (complementary log-log transformation). They have the following properties:

1) Probit: $\pi_i \; = \; \Phi \, (\eta_i); \;\; \eta_i \; = \; \Phi^{-1} \, (\pi_i);$
2) Logit: $\pi_{\;i} \; = \; F \, (\eta_i) \; = \; \frac{e^{\eta_i}}{1 + \, e^{\eta_i}}; \;\; \eta_i \; = \; F^{-1} \, (\pi_i) = log \, \frac{\pi_i}{1 - \, \pi_i};$
3) Complementary log-log transformation: $\pi_{\;i} \; = \; e^{-e^{-\eta_i}}; \;\; \eta_i \; = \log(-\log(1 - \pi_{\;i})).$

The author highlights, “The complementary log-log transformation has a direct interpretation in terms of hazard ratios, and thus has practical implications in terms of hazard models…”. Also, Oosterbaan (1994) in their “Chapter 6: Frequency and regression analysis of hydrologic data” in Ritzema H. P. “Drainage Principle and Applications, Publication 16” claims, “they [hydrologists] frequently use the Gumbel distribution to find annual or monthly maxima of floods or to find rainfalls of short duration”. The author also presents an example of finding a maximum of some hydrological process using Gumbel distribution paper. In fact, this paper has complementary log-log transformation of the considered function as a scale which yields a necessary linear relationship with a variable that denotes a wanted maximum.

As we discussed at the beginning of this chapter, in step two the advanced time-series techniques are used with respect our variable of choice from step one – rainfall precipitation. Enders (2014) in his book “Applied econometric time series” points out that time-series could be decomposed into four main components: trend, seasonal, cyclical, and irregular (or stochastic) component. The first two of them are associated with long-run dynamics of the series. The author of the book highlights that autoregressive process of order one or AR (1) with linear time trend and seasonal component is the simplest and the most reliable way to capture the major part of the long-run mean for a given data and/or model. Rainfall precipitation data in our case is in annual format and not a subject to the seasonal component. Thus, our variable of choice is decomposed via AR (1) process with linear time trend. The rain variable and its residuals from the estimation mentioned above are tested for the breakpoint. To do so, Zivot-Andrews test is used as one of the widely used tests in the literature.

Finally, the two steps are combined to make a reliable forecast of the flooding events in Atlantic Canada through changes in the time path of the rainfall precipitation.

Chapter 4. Estimations and results

4.1 Model specification and probability of floods in Atlantic Canada

The flood forecasting system should be transparent and reliable. As previously discussed, I am looking for a straightforward model that does not require a board of specialists in various fields and thus, it could be operated with minimum human capital costs.

Most of the literature suggests using time-series analysis for extreme weather events estimation. I am going to follow this trend as it is proved to be trustworthy and intelligible theory for my purposes.

There is no common opinion among researchers with regards to floods determinants. As we indicated above, various literature proposes different variables such as: annual precipitation, annual snow, annual rain, deviation of average precipitations from their historical trends, number of series of consecutive days with heavy precipitation; annual average temperature, deviation of annual average temperature from its trend, deviation of the average temperature at the month of the event; annual discharge level mean, annual maximum discharge; mean sea level. As I am interested in performing analysis for the local scope, I was limited by data and unable to collect it for some of the variables above.

The dependent variable in my model in step one is flood dummy takes on value of one when at least one flood occurred in a given year and zero otherwise. Based on the data from Canadian Disaster Database (CDD), we would mostly investigate the impact of inland floods. As we use data from CDD, definition of flood in our case coincides with Burina (2017) presented in the introduction. As independent variables I included annual rain, annual snow, annual average temperature, annual maximum discharge, mean sea level and year of observation. In addition, one of the economic variables – income – was incorporated into specification to address possible relationship between floods and economic performance over time. I would like to check whether change in real disposable income per capita influences probability of floods. It might be the case that with more per capita income people become wealthier; thus, they build more infrastructure and become more exposed to floods in context of definition from CDD.

I expect to observe positive relationship between floods and rain, snow, discharge and sea level variables as these variables might be considered as pre-determinants of flood. While I suppose to find a negative relationship between floods and temperature variable, because higher annual average temperature usually leads to drier climate patterns. All variables, their descriptions and sources are presented below. Information on location of gauges and descriptive statistics for the relevant variables is also presented.

Variables description:

year – year of observation;

dummy_flood – binary variable (1 – flood, 0 – no flood);

rain – annual amount of rain in mm;

snow – annual amount of snow in cm;

temperature – annual average mean temperature in ^{0}C;

discharge – annual maximum river discharge in m^3/s;

sea – mean sea level in mm;

income – real disposable per capita income in dollars 2002;

log"varname" – natural logarithm of the corresponding variable.

Table 2: Source of data

Variable	Source
dummy_flood	Canadian Disaster Database; NB Flood History Database
rain	Environment and Climate Change Canada
snow	Environment and Climate Change Canada
temperature	Environment and Climate Change Canada
discharge	Environment Canada
sea	Permanent Service for Mean Sea Level (PSMSL)
income	Statistics of Canada

Table 3: Location of gauge

Variable	Location
rain	UNB, Fredericton
snow	UNB, Fredericton
temperature	UNB, Fredericton
discharge	Saint John River, Grand Falls
sea	Saint John

Table 4: Descriptive statistics

Variable	Number of obs.	Mean	Min	Max
year	86	1972.5	1930	2015
dummy_flood	86	0.372093	0	1
rain	86	854.7837	474	1266.1
snow	86	263.7279	75.7	468.8
temperature	86	5.443023	3.8	7.4
discharge	86	3328.081	915	7500
sea	86	7004.733	6861	7209
income	84	10867.59	2225.7	22338.7

Therefore, all three models were estimated in Stata in levels and in logarithms. The latter was done to test for possible non-linearities in the model. The basic model was estimated in probit, logit and complementary log-log transformation framework. The main idea here is to choose the best specification among these three models and to proceed with it. The following tables 2 and 3 present our results.

Table 5: Estimation results in levels for all variables (standard errors in parentheses)

Variable	Probit		Logit		Comp. log-log	
	Coef	**z-value**	**Coef**	**z-value**	**Coef**	**z-value**
rain	0.0039005 (*0.0013269*)	2.94	0.0072818 (*0.0024786*)	2.94	0.0059157 (*0.0018527*)	3.19
snow	-0.0014373 (*0.0029*)	-0.48	-0.0022962 (*0.0052258*)	-0.44	-0.0012856 (*0.0034043*)	-0.38
temperature	-0.5455037 (*0.2905437*)	-1.88	-0.9085371 (*0.5039222*)	-1.80	-0.7141765 (*0.3918479*)	-1.82
discharge	0.0005595 (*0.000164*)	3.41	0.0010066 (*0.0003074*)	3.27	0.0007534 (*0.0002106*)	3.58
sea	0.0099308 (*0.0047127*)	2.11	0.0162102 (*0.0082674*)	1.96	0.0105088 (*0.0054879*)	1.91
income	-0.0000387 (*0.0000452*)	-0.86	-0.0000576 (*0.0000766*)	-0.75	-0.0000344 (*0.0000544*)	-0.63
constant	-71.5155 (*32.08383*)	-2.23	-117.8992 (*56.36849*)	-2.09	-77.761 (*37.51056*)	-2.07
	AIC = 1.073097 BIC = -265.0328 Pseudo R^2 = 0.3046		AIC = 1.066484 BIC = -265.5882 Pseudo R^2 = 0.3097		AIC = 1.064656 BIC = -265.7418	

Table 6: Estimation results in logs for all variables (standard errors in parentheses)

Variable	Probit		Logit		Comp. log-log	
	Coef	**z-value**	**Coef**	**z-value**	**Coef**	**z-value**
lograin	2.974064 (*1.078251*)	2.76	5.647637 (*2.0272*)	2.79	4.622046 (*1.55687*)	2.97
logsnow	0.1215084 (*0.7289998*)	0.17	0.2589357 (*1.247278*)	0.21	0.2334223 (*0.8859702*)	0.26
logtemperature	-2.192946 (*1.51941*)	-1.44	-3.611766 (*2.60924*)	-1.38	-2.502822 (*1.957763*)	-1.28
logdischarge	1.677533 (*0.5369672*)	3.12	3.157656 (*1.030367*)	3.06	2.553728 (*0.8015637*)	3.19
logsea	64.23677 (*34.55837*)	1.86	105.5395 (*60.1431*)	1.75	69.902 (*41.23289*)	1.70
logincome	-0.3571976 (*0.440307*)	-0.81	-0.5518619 (*0.7441262*)	-0.74	-0.3933789 (*0.540855*)	-0.73
constant	-596.5655 (*302.1778*)	-1.97	-989.2509 (*527.3135*)	-1.88	-665.3338 (*361.304*)	-1.84
	AIC = 1.113163 BIC = -261.6672		AIC = 1.100789 BIC = -262.7066		AIC = 1.092315 BIC = -263.4184	

	Pseudo R^2 = 0.2739	Pseudo R^2 = 0.2834	

Stata does not report pseudo R^2 (analogue of R^2 for GLM) for complementary log-log transformation, but it has slightly better fit in comparison with logit and probit. But in general Stata manual suggests "Because this statistic [pseudo R^2] does not mean what R^2 means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic with great caution". Hosmer and Lemershow, in their textbook "Applied Logistic Regression" explain that pseudo R^2 is not useful to report alone but might be reasonable statistic to evaluate competing model. Nevertheless, smaller outcome of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) show the model with better relative power. Based on this data set and estimations, we can state that our models have the following ranking:

1) Complementary log-log transformation;
2) Logit;
3) Probit.

But as we can see not all the variables are statistically significant. According to Stata manual, variable is statistically significant if |z-value| ≥ 1.96. Based on this definition, for example, snow and income are far from being significant. Based on our data, we did not find any statistical evidences for the relationship between proxy of economic performance – income – and probability of flood occurrence in Atlantic Canada. Results with only significant variables are presented below in tables 4 and 5.

Table 7: Estimation results in levels for statistically significant variables (standard errors in parentheses)

Variable	**Probit**		**Logit**		**Comp. log-log**	
	Coef	**z-value**	**Coef**	**z-value**	**Coef**	**z-value**
rain	0.0040588 (*0.0013048*)	3.11	0.0074721 (*0.0024297*)	3.08	0.0059702 (*0.0018086*)	3.30

temperature	-0.5518679 (*0.2508666*)	-2.20	-0.9361392 (*0.4460224*)	-2.10	-0.7155042 (*0.3144972*)	-2.28
discharge	0.0005245 (*0.0001613*)	3.25	0.0009487 (*0.0003008*)	3.15	0.0007123 (*0.0002085*)	3.42
sea	0.0069015 (*0.0027459*)	2.51	0.0116321 (*0.0047793*)	2.43	0.0077618 (*0.0034015*)	2.28
constant	-51.04924 (*18.98468*)	-2.69	-86.82574 (*33.27439*)	-2.61	-59.09007 (*23.82607*)	-2.48
	AIC = 1.034039 BIC = -281.8748 Pseudo R^2 = 0.3048		AIC = 1.027229 BIC = -282.4604 Pseudo R^2 = 0.3099		AIC = 1.023099 BIC = -282.8156	

Table 8: Estimation results in logs for statistically significant variables (standard errors in parentheses)

Variable	**Probit**		**Logit**		**Comp. log-log**	
	Coef	**z-value**	**Coef**	**z-value**	**Coef**	**z-value**
lograin	3.066071 (*1.058162*)	2.90	5.734965 (*1.983421*)	2.89	4.579901 (*1.507495*)	3.04
logtemperature	-2.688272 (*1.320021*)	-2.04	-4.550076 (*2.314561*)	-1.97	-3.267476 (*1.566408*)	-2.09
logdischarge	1.615926 (*0.5281205*)	3.06	3.040746 (*1.003948*)	3.03	2.420088 (*0.772697*)	3.13
logsea	46.28019 (*18.93879*)	2.44	78.19842 (*33.07324*)	2.36	50.97309 (*22.92133*)	2.22
constant	-439.4161 (*168.1415*)	-2.61	-748.7708 (*295.4921*)	-2.53	-497.3237 (*204.2823*)	-2.43
	AIC = 1.069614 BIC = -278.8153 Pseudo R^2 = 0.2778		AIC = 1.058753 BIC = -279.7494 Pseudo R^2 = 0.2861		AIC = 1.05196 BIC = -280.3335	

These estimations show that there are 4 statistically significant variables: rain, temperature, discharge, sea and corresponding variables in logarithms. Moreover, rain appears to be the most significant hydro-climatic variable in complementary log-log transformation model. Also, we can conclude that relative ranking of models has not change.

I would like to point out that signs of the coefficients are seemed to follow our expectations.

Based on the estimation above, there is no substantial difference between levels and logarithms in terms of statistical significance of coefficients among all three models. Hence, I decided to continue further analysis in levels only.

4.2 Dynamics of rainfall in Atlantic Canada

According to the Fourth Assessment Report of Intergovernmental Panel on Climate Change (IPCC) temperature and precipitation changes are the most "obvious and easily measured changes in climate". It is also clear from our estimations that their changes have a significant impact on floods frequency and severity. To build more precise and reliable model, rain variable would be "decomposed" further into its long-run mean that characterizes frequency of rainfall and residuals that would be tested for some variance patterns that correspond to intensity or magnitude of the rainfalls. Variable rain was chosen for this estimation as it is one of the major determinants of climate change and floods based on the estimations from step one. Moreover, rainfall variable has a twofold nature – it is simultaneously hydrological and climatic variable. It also has detailed data not only on annual but monthly and daily basis. The latter might be useful for further investigations on this topic.

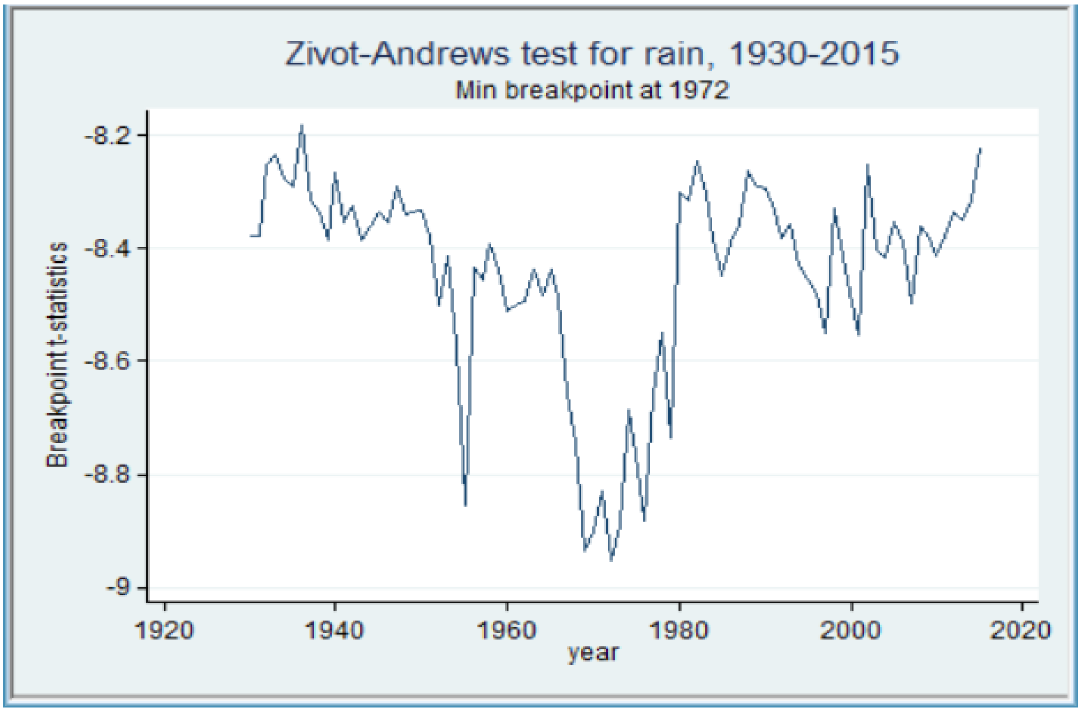
As discussed in the methodology, the rainfall data is estimated using AR (1) model with linear time trend. The results are presented below.

Table 9: Estimation results for rain: AR (1) with trend

Variable	Coefficient	z-value
AR-1	0.831885	0.76
trend	0.187374	0.27
constant	777.3404	7.92

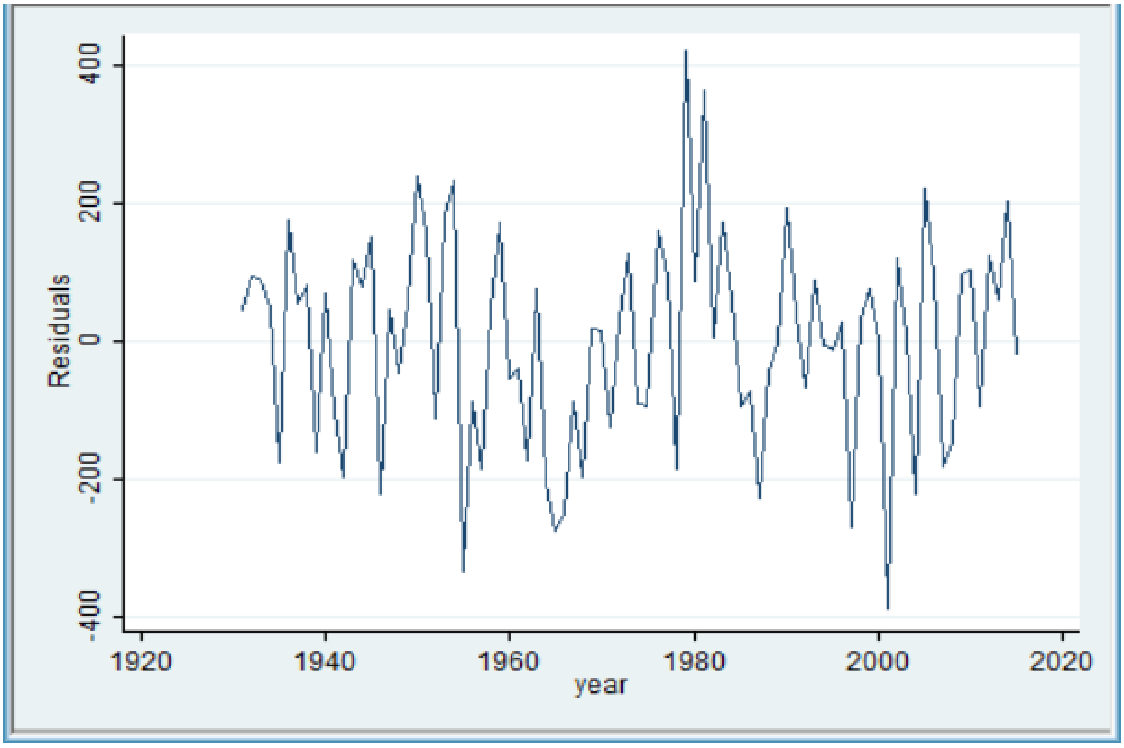
As we can see, both trend and AR part are not very statistically significant. That's why the rain data was further tested for a structural break. In general, change in a dynamic process can be of different nature: (i) change in the process's mean, (ii) change in the process's variance, (iii) change in both, the process's mean and variance. I have used Zivot-Andrews test for structural break which is allowing for break in both intercept and trend. In a few words, this test has its own breakpoint statistics that is being calculated for each observation and then the observation with the most negative value of statistics is reported to be the most feasible candidate for a breakpoint. Stata also supports a graph option to visualize the outcome.

Figure 1: Zivot-Andrews test for the rain variable



As we can see from the graph, Zivot-Andrews test states the most viable candidate for a breakpoint year is 1972. I have tried different ARMA structures with trend before and after this year, but none of them provides a set of statistically significant coefficients. As a result, I decided to stay with AR (1) structure with trend and focus on residuals from this regression to find some patterns in magnitude of the rain variable. The graph for residuals could be found below.

Figure 2: Graph for residuals



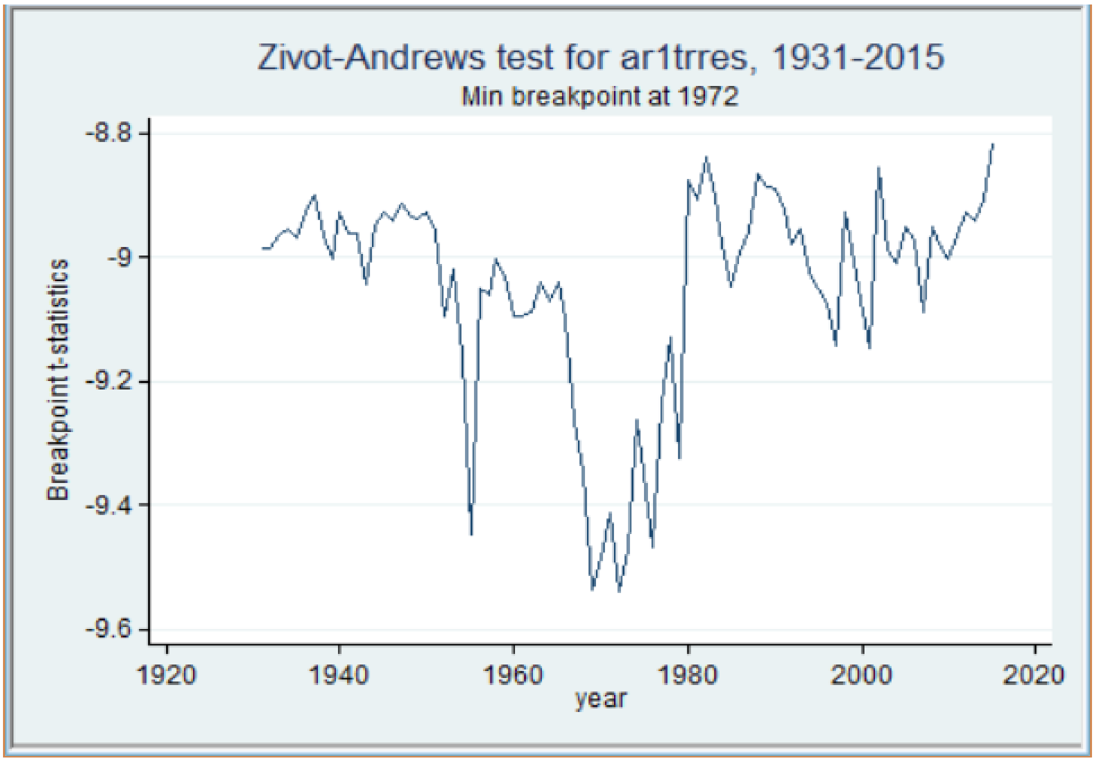
Residuals are seemed to have complex variance structure with possible break in it. That's why I would like to use ARGH/GARCH models. The main idea here is to find good fit for the whole sample and test for a structural break. The moderate overall fit for the residuals is shown with ARCH (1) GARCH-3 structure.

Table 10: Estimation results for the residuals: ARCH (1) GARCH-3

Variable	Coefficient	z-value
ARCH-1	- 0.0905263	- 1.47
GARCH-3	- 0.7378028	- 2.54
constant	42058.92	4.58

Residuals were also tested for structural break.

Figure 3: Zivot-Andrews test for the residuals



Zivot-Andrews test reports the same year as a potential breakpoint – 1972. But in the case of residuals, statistical difference was found with respect to significance of the coefficients in the previous ARCH (1) GARCH-3 structure.

Table 11: Estimation results for the residuals: ARCH (1) GARCH-3 before 1972

Variable	Coefficient	z-value
ARCH-1	0.2239242	0.89
GARCH-3	0.4906698	0.61
constant	6495.686	0.37

Table 12: Estimation results for the residuals: ARCH (1) GARCH-3 after 1972

Variable	Coefficient	z-value
ARCH-1	- 0.1258936	- 2.31
GARCH-3	- 0.8291315	- 4.94
constant	49122.67	3.56

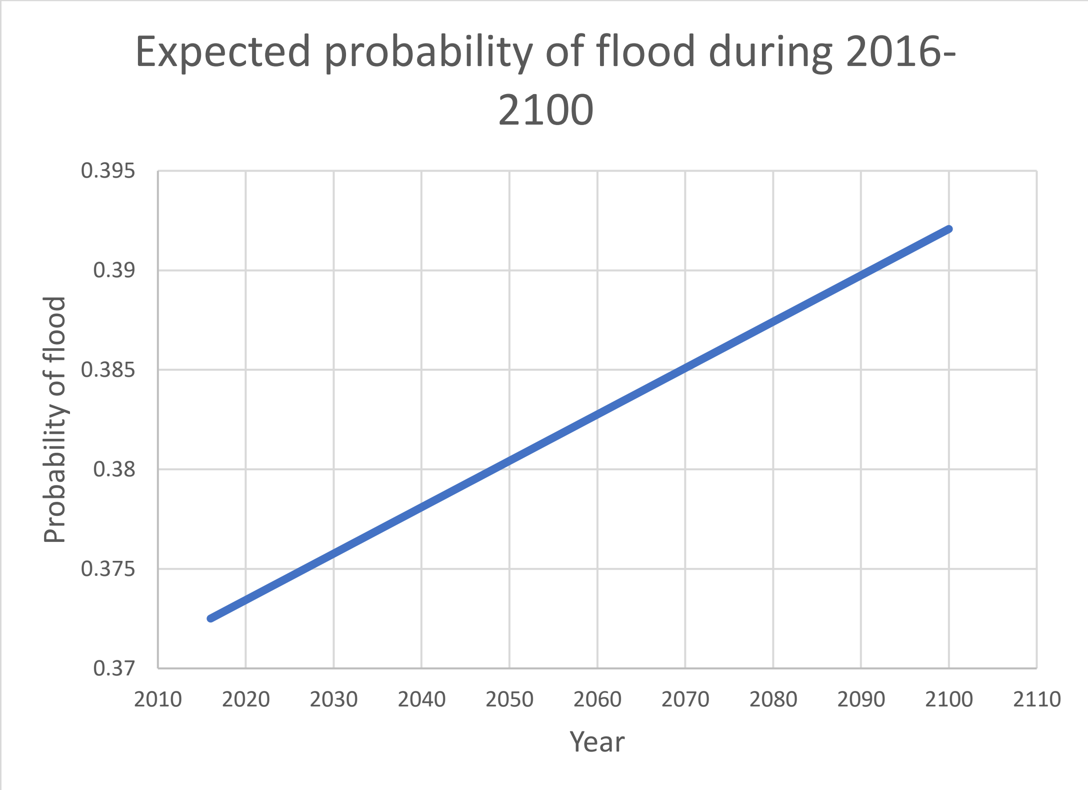
As we can see from two estimations above, coefficients of the model are insignificant before 1972 and become highly significant after 1972. That’s why it is reasonable to suggest that 1972 year is a breakpoint in our dataset. This breakpoint suggests that variance or magnitude of the rain variable changes over time.

Based on the data we use, I can state that rain variable does not have highly significant trend, but there is exist increase in magnitude with respect to rain patterns in Atlantic Canada. Since rain variable is one of the major determinants of flood in our model and rain is positively correlated with flood occurrence, I can claim that flood behavior follows the same patterns as rain does.

4.3 Probability of floods and climate change

To show how our methodology works, I would use not very statistically significant trend for the rain variable that is equal to 0.187374. First, I found mean of the estimated probability of flood according to our estimation based on complementary log-log transformation model. This mean is approximately equal to 0.372505 and I denote it as initial probability and assign this value to 2016 – the following year after the last year in our sample. Then I intend to find rate of change of the probability of flood occurrence in the future. To perform this, I multiply trend coefficient in the rain estimation by the rain coefficient in the basic model of the flood probability estimation in step one. Since our initial estimation is done using link function – complementary log-log transformation, it was necessary to convert rain coefficient back to marginal effect. It was done using Stata postestimation command *margins*. When all covariates are at observed values, the marginal effect of rain variable is equal to 0.0012437. It means that the probability of occurring flood in the given year increases by the number above with every additional millimeter of rain per year. So, expected change in the probability of flood occurrence per year is equal to: $0.0012437 \times 0.187374$, which is approximately equal to 0.000233 or 0.0233%. This is *annual change of probability of flood* in Atlantic Canada. The resulting graph that presents our forecast from 2016 to 2100 is shown below.

Figure 4: Expected probability of flood by the end of the current century



Coupled with other studies, it is possible to set up policy recommendations. For example, Burina (2017) suggests expected average loss from floods is equal to $9.5 millions annually. Based on our findings, we could state expected loss from floods in Atlantic Canada will increase by 9,500,000×0.02 = $190,000 on annual basis by the end of the current century.

Chapter 5. Conclusions, limitations and future work

Extreme weather events are one of the most destructive and damaging consequences of climate change. This study shows that floods prevails among various large weather events in Atlantic Canada. This paper tests the hypothesis that climate change leads to an increase in frequency and magnitude of floods. To show this, a two-step procedure was established. In step one, we estimate probability of floods as a function of climate and hydrological variables based on complementary log-log transformation model. We found a positive relationship between probability and rainfall – major climate and hydrological variable simultaneously. In step two, dynamics of rainfall time series was analyzed and decomposed to find exact time path of it. It is shown in the study that rainfall magnitude increases over time; change in rainfall variance is considered as demonstration of climate change in Atlantic region. Finally, two steps were combined to make a framework for probability forecasting. If we follow our trend point estimate from step two, the results show that probability of floods is going to increase by two percentage points by the end of this century due to ongoing climate change. Since climate change in this study is viewed via change in the dynamics of rainfall, another important conclusion of this study is: climate change in Atlantic Canada can be dated back to 1972. This conclusion is in accord with earlier conducted studies in the region specifically the ones done in the Department of Economics at the University of New Brunswick (Fredericton, Canada).

However, there are two main limitations in this study. First, I focused mostly on climate and hydrological variables as floods predictors. It is possible that some other economic or social variables would be reasonable to include in our model.

The second limitation is that we have used aggregated rainfall variable – in annual format. As we can see from our estimations, linear trend of the rain variable is not very significant. It might be explained by the fact that rain patterns could be sharply different within the province during some periods of time within a year. I assume that using the monthly or ideally daily data could improve the situation towards higher significance.

Finally, future research would be essential to address our limitations. One might add some additional variables as flood predictors and test their relevance to the model. For example, it could be average temperature in the month of the event or deviation of annual average temperature from its trend. More importantly, less aggregated data, especially for rainfall series, could be used to provide even stronger results. As stated above, monthly or daily data should provide more reliable outcome of the model.

Bibliography

Intergovernmental Panel on Climate Change. 2013. Climate Change 2013: The Physical Science Basis. *IPCC*. Accessed January 4, 2018. http://www.ipcc.ch/report/ar5/wg1/.

Public Safety Canada. n.d. Canadian Disaster Database. *Pubic Safety Canada*. Accessed on September 28, 2017. http://cdd.publicsafety.gc.ca/rslts-eng.aspx?cultureCode=en-Ca&boundingBox=&provinces=4&eventTypes=%27FL%27&eventStartDate=&injured=&evacuated=&totalCost=&dead=&normalizedCostYear=1&dynamic=false.

Huber, Daniel G., and Jay Gulledge. 2011. "Extreme Weather and Climate Change: Understanding the Link and Managing the Risk". *Centre for Climate and Energy Solutions*. Accessed on October 23, 2017. https://www.c2es.org/site/assets/uploads/2011/12/white-paper-extreme-weather-climate-change-understanding-link-managing-risk.pdf.

Hlavcova, K., S. Kohnova, R. Kubes, J. Szolgay, and M. Zvolensky. 2005. "An Empirical Method for Estimating Future Flood Risks for Flood Warnings". *Hydrology and Earth System Sciences* 9 (4): 431–448. https://www.hydrol-earth-syst-sci.net/9/431/2005/hess-9-431-2005.pdf.

Vaze, J., P. Jordan, R. Beecham, A. Frost, G. Summerell. 2012. "Guidelines for Rainfall-runoff Modelling: Towards Best Practice Model Application". *eWater Cooperative Research Centre*. Accessed on October 24, 2017. https://ewater.org.au/uploads/files/eWater-Modelling-Guidelines-RRM-(v1-Mar-2012).pdf.

Zhijia, Li, Wang Lili, Bao Hongjun, Song Yu, Yu Zhongbo. 2008. "Rainfall-runoff Simulation and Flood Forecasting for Huaihe Basin". *Science Direct* 1 (3): 24–35. https://www.sciencedirect.com/science/article/pii/S1674237015300247

Calver, Ann, R.Lamb and S. E. Morris. 1999. "River Flood Frequency Estimation Using Continuous Runoff Modelling". *Proceedings of the Institution of Civil Engineers – Water, Maritime and Energy* 136 (4): 225–234. https://www.icevirtuallibrary.com/doi/10.1680/iwtme.1999.31986.

Cameron, David, Keith Beven and Pamela Naden. 2000. "Flood Frequency Estimation by Continuous Simulation under Climate Change (with Uncertainty)". *Hydrology and Earth System Sciences* 4 (3): 393–405. https://www.hydrol-earth-syst-sci.net/4/393/2000/hess-4-393-2000.pdf.

Vaze, Jai, F. H. S. Chiew, J. M. Perraud, N. Niney, D. Post, J. Teng, B. Wang, J. Lerat, & M. Goswami. 2011. "Rainfall-Runoff Modelling Across Southeast Australia: Datasets, Models and Results". *Australasian Journal of Water Resources* 14 (2): 101–116. https://www.tandfonline.com/doi/abs/10.1080/13241583.2011.11465379.

Madsen, H. 2000. "Automatic Calibration of a Conceptual Rainfall-Runoff Model Using Multiple Objectives". *Journal of Hydrology* 235 (3-4): 276–288. https://www.sciencedirect.com/science/article/pii/S0022169400002791#!

Kundzewicz, Z. W et al. 2013. "Flood Risk and Climate Change: Global and Regional Perspectives". *Hydrological Sciences Journal* 59 (1): 1-28.

Paeth, Heiko, Andreas H. Fink, Susan Pohle, Felix Keis, Hermann Machel, Cyrus Samimi. 2010. "Meteorological Characteristics and Potential Causes of the 2007 Flood in sub-Saharan Africa". *International Journal of Climatology* 31

(13): 1908–1926. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.2199.

Casse, C., M. Gosset, C. Peugeot, V. Pedinotti, A. Boone, B. A. Tanimoun, B. Decharme. 2015. "Potential of Satelite Rainfall Products to Predict Niger River Flood Events in Niamey". *Atmospheric Research* 163: 162–176. https://www.sciencedirect.com/science/article/pii/S0169809515000277?via%3Dihub.

Sarr, Mamadou Adama, Malicki Zorome, Ousmane Seidou, Christopher Robin Bryant, Philippe Gachon. 2013. "Recent Trends in Selected Extreme Precipitation Indices in Senegal – A Changepoint Approach". *Journal of Hydrology* 505: 326–334. https://www.sciencedirect.com/science/article/pii/S0022169413006859?via%3Dihub.

Hasanah, Yulianti, Marizsa Herlina, Hilda Zaikarina. 2013. "Flood Prediction using Transfer Function Model of Rainfall and Water Discharge Approach in Katulampa Dam". *Science Direct* 17: 317–326. https://www.sciencedirect.com/science/article/pii/S1878029613000467.

Aich, Valentin, Bakary Kone, Fred F. Hattermann, and Eva N. Paton. 2016. "Time Series Analysis of Floods across the Niger River Basin". *Water 2016* 8 (4): 165. doi: 10.3390/w8040165. http://www.mdpi.com/2073-4441/8/4/165.

Mohr S., M. Kunz, K. Keuler. 2015. "Development and Application of a Logistic Model to Estimate the Past and Future Hail Potential in Germany". *Journal of Geophysical Research: Atmospheres* 120 (9): 3939–3956. http://onlinelibrary.wiley.com/doi/10.1002/2014JD022959/full.

Mouri, Goro, Daigo Minoshima, Valentin Golosov, Sergey Chalov, Shinta Seto, Kei Yoshimura, Shinichiro Nakamura, Taikan Oki. 2013. "Probability Assessment of Flood and Sediment Disaster in Japan Using the Total Runoff-Integrating Pathways Model". *Science Direct* 3: 31–43. https://www.sciencedirect.com/science/article/pii/S2212420912000398.

Burina, Iuliia. 2017. "Economic consequences of floods in Maritime provinces". *MA Report in Economics*. University of New Brunswick.

Davidson, Russell and James G. MacKinnon. 2004. *Econometric Theory and Methods*. The United States of America: Oxford University Press.

Rodriguez, German. n.d. "Generalized Linear Models: Other Choices of Link". *Princeton University*. Accessed on January 11, 2018. http://data.princeton.edu/wws509/notes/c3s7.html.

Oosterbaan, R. J. 1994. "Frequency and Regression Analysis in Hydrologic Data". Chapter 6 in *Drainage Principles and Applications*, edited by H. P. Ritzema. Wageningen, The Netherlands: revised by International Institute for Land Reclamation and Improvement (ILRI). https://www.waterlog.info/pdf/freqtxt.pdf.

Stata Manual.

Environment and Local Government NB. n.d. *Flood History Database*. Accessed on January 8, 2018. http://www.elgegl.gnb.ca/0001/en/Flood/Search.

Environment and Climate Change Canada. n.d. *Historical Climate Data*. Accessed on November 17, 2017. http://climate.weather.gc.ca/index_e.html.

Statistics Canada. n.d. *Data on Long-run Provincial and Territorial Performance*. Accessed on December 3, 2017. http://www5.statcan.gc.ca/cansim/a47.

Enders, Walter. 2014. *Applied Econometric Time Series*, 4th edition. The United States of America: Wiley.

Curriculum Vitae

Candidate's full name: Stanislav Hetalo.

Universities attended (with dates and degrees obtained): Taras Shevchenko National University of Kyiv, Ukraine, 2012 – 2016, Bachelor of System Analysis.

Publications: None.

Conference Presentations: 1) Stanislav Hetalo, 2018: "Large Weather Events and Climate Change: Floods in Atlantic Canada", The Graduate Research Conference. University of New Brunswick; 2) Stanislav Hetalo, 2012: "Numerical constructions. The sum of cubes and squares", The International Conference of Young Scientists 2012. Radboud University Nijmegen.