UPTEC W08 001

Examensarbete 30 hp
Januari 2008

SLU

Sveriges
lantbruksuniversitet

Modelling the effects of catchment properties on DOC fluxes in the MRW, Ontario, Canada

Modellering av effekterna från egenskaper hos avrinningsområdet på DOC flödet i MRW

Magdalena Nyberg

ABSTRACT

Modelling the effects of catchment properties on DOC fluxes in the Muskoka River Watershed, Ontario, Canada

Magdalena Nyberg

Dissolved organic carbon, DOC, has major effects on ecosystems as it influences soil formation, forms complexes with metals and nutrients affecting their flux and bioavailability. It also reacts with chlorine from water treatments, forming THM a carcinogenic substance. The effects of climate change have been linked to release of greenhouse gases. As CO_2 is a major greenhouse gas all parts of the global carbon cycle have become a research interest.

This study is a continuation of the search to find simple, black box, mass balance models that successfully estimate stream DOC concentrations from catchment properties. Using GIS, Geographic Information Systems, 26 parameters relating to properties of the 20 subcatchments were investigated leading to the identification of eight models with one to three parameters. The models utilized six of the available parameters. Choosing fifteen different, unique subcatchments for 10 000 runs where those fifteen were used for calibration and the remaining five subcatchments for validation, mean coefficients were obtained. These were used in a sensitivity analysis, and based on the result three models were chosen. Model M1 only contained the average slope of the catchment, M3; building on the framework of M1, also included percentage wetland and M8 added drainage density as a third parameter.

The three chosen models, as well as a fourth model from 1997, derived from the same area and containing peat (wetland) percentage as the only parameter, were then tested on the Muskoka River Watershed in Ontario. Each model was linked to the lake DOC Model (LDM) to connect all the 859 lakes in the watershed and to gain an estimate of the lake DOC concentrations. The Lake DOC Model was also optimized twice for each model, once for all the 237 lakes with measured values of DOC and then for only those 117 lakes that were headwater lakes. Optimization was made to minimize the average absolute deviations of the estimated values.

The results were that M1 explained about 45 % of the DOC concentration in the lakes, M3 46-47 %, M8 47-49 % and the older 1997 model 44-53 %. The mean of the estimated DOC from the three derived models and the mean from the three models and the 1997 model, explained 47 % and 50-54 % respectively. That means that the best result was that of the mean estimate of all four models.

Keywords: Dissolved organic carbon, wetlands, fluxes, GIS, mass balance, Muskoka River Watershed, Dorset study, climate change, biogeochemistry

Department of Environmental Assessment, SLU, Vallvägen 3, SE-756 51 Uppsala, Sweden
ISSN 1401-5765

REFERAT

Modellering av hur egenskaper hos avrinningsområdet påverkar DOC flödet i flodområdet Muskoka, Ontario, Kanada

Magdalena Nyberg

Löst organisk kol, DOC, har en avgörande effekt på olika ekosystem då det påverkar formationen av jordarter och bildar komplex med metaller och näringsämnen, vilket påverkar deras flöden och biologiska tillgänglighet. Det reagerar också med klor från vattenrening, därmed bildas THM, en cancerogen substans. Klimatförändringar har kopplats till frigörandet av växthusgaser. Eftersom CO_2 är en avgörande växthusgas så blir alla delar av den globala kolcykeln intressanta ur forskningssynpunkt.

Denna studie är en fortsättning på tidigare studier som sökt efter enkla, black box-, massbalans modeller som framgångsrikt kan uppskatta flodkoncentrationen av DOC med egenskaper hos avrinningsområdet. Med GIS (Geographic Information Systems), erhölls 26 parametrar beskrivande egenskaper hos de 20 delavrinningsområdena vilka undersöktes och åtta, en till tre parameters modeller som utnyttjade totalt sex av de tillgängliga parametrarna, identifierades. Genom att välja femton olika, unika delavrinningsområden 10 000 gånger och varje gång kalibrera med dessa femton och validera mot de resterande fem delavrinningsområdena, förvärvades medelkoefficienter. Dessa användes i modeller för en känslighetsanalys, och baserat på resultatet valdes tre modeller. Modell M1 innehåll endast medellutning i avrinningsområdet, M3, som byggde på M1´s stomme, innehöll även procent våtmark och M8 adderade också dräneringsdensitet som en tredje parameter.

De tre modellerna, liksom en fjärde modell från 1997, som erhållits från samma områden i Dorset och innehåller torv (våtmarks) procent som enda parameter, testades på flodområdet Muskoka i Ontario. Varje modell länkades till Sjö DOC Modellen (LDM) för att kunna koppla samman de 859 sjöarna i avrinningsområdet och få en uppskattning av sjökoncentrationerna av DOC. Sjö DOC Modellen optimerades också två gånger för varje modell, för alla 237 sjöar med mätvärden och för de 117 sjöar som var källsjöar (sjöar av första ordningen). Optimering skedde genom att försöka minimera absolutvärden av medelavvikelsen av de uppskattade värdena.

Resultatet var att M1 förklarade 45 % av koncentrationen av DOC i sjöarna, M3 46-47 %, M8 47-49 % och modellen från 1997 44-53 %. Medel av de uppskattade DOC värdena baserade på de tre framtagna modellerna eller alla fyra förklarade 47 % och 50-54 % respektive. Det gör att det bästa resultat kom från medeluppskattningar av alla fyra modellerna.

Nyckelord: Löst organisk kol, våtmarker, flöden, GIS, massbalans, flodområdet Muskoka, Dorset studien, klimatförändring, biogeokemi

Institutionen för Miljöanalys, Sveriges Lantbruksuniveristet, Vallvägen 3, 756 51 Uppsala
ISSN 1401-5765

PREFACE

This thesis project was completed at Trent University in Peterborough, Ontario, Canada but is a part of the M. Sc. in Aquatic and Environmental Engineering at Uppsala University, covering 20 Swedish academic points (30 ECT´s). The supervisor in Ontario was Peter Dillon at the Department of Environmental Science, Trent University. The Subject Reviewer in Sweden was Kevin Bishop at The Swedish University for Agricultural Science, Dept. of Environmental Assessment, SLU (The Swedish University of Agricultural Sciences).

This Thesis project is a part of an NSERC Strategic Grant on Modelling DOM in the Great Lakes Basin.

ACKNOWLEDGEMENTS

My supervisor Peter Dillon has my greatest thanks for taking me on, after not even meeting me first, which made this project possible. I thank Eavan O'Connor for helping me with the original model, which she already used, and Madeline that made sure I ended up in Peterborough when I first arrived. Heather Broadbent that made sure all papers were filled out and that I was on top of things. Helen Baulch, Stephen Oni and others for, thanks for the company during the sometimes long days at Trent, knowing you are not alone helps.

I thank Joe F for teaching me how to perform the Brien´s test, with personal help and instructions, and for answering other questions that I had at times. James Ouellette, Marta, Jessica, Clinton and others for their GIS help. GIS layers were supplied under a licence agreement by Members of the Ontario Geospatial Data Exchange and I thank them for making it possible to complete this work. Some data was also supplied by the regional MNR/FRI offices. I thank them very much, especially Dave Miles for supplying me with data as well as phone numbers. I thank Clint also for VBA code for ArcGIS and Marc Cadieux for his help with VBA code in Excel.

I also thank my subject reviewer, Kevin Bishop, for being a solid support overseas, and some other people at SLU that were helpful in supplying me with articles and statistical assistance.

Colin, Brent and Dylan for sharing a living place with me.

For all others around me that have not been specifically mentioned but have been equally important, like fellow hikers and neighbours, I have not forgotten and I thank you.

UPTEC W08 001, ISSN 1401-5765
Printed at the Department of Earth Sciences, Geotryckeriet, Uppsala University, Uppsala 2008

POPULÄRVETENSKAPLIG SAMMANFATTNING

Modellering av hur egenskaper hos avrinningsområdet påverkar DOC flödet i flodområdet Muskoka, Ontario, Kanada

Magdalena Nyberg

Löst organiskt kol i mark, floder och sjöar är en del av den globala kolcykeln och som sådan kan ekosystem bidra till eller fånga upp koldioxid i atmosfären. Med den ökande växthuseffekten och de efterföljande klimatförändringarna som delvis beror av CO_2 kan dessa flöden förändras vilket påverkar växtlighet, men kan också öka eller minska koncentrationen i atmosfären.

Kol samverkar också med andra komponenter i mark och vatten såsom metaller och näringsämnen. Dessa ämnens flöde och biologiska tillgänglighet påverkas av de komplex som bildas. Kol reagerar också med klor som är en del i rening av vatten i reningsverk och vid denna reaktion bildas ett cancerogent ämne som förkortas THM.

Vid Trent Universitet i Peterborough, Ontario, Kanada pågår ett stort projekt som försöker förklara flödet av DOC och sedan utifrån detta kunna förbättra uppskattningen även av andra ämnens flöden. Bland annat har man tittat på kopplingen till flödet av kvicksilver och därmed också halten av kvicksilver i fisk i sjöar. Man hoppas kunna få fram (minst) en bra modell för flödet av dessa ämnen med indata från tillgängliga GIS (Geografiska Informations System) data. Denna/dessa modeller hoppas man sedan kunna använda för att uppskatta koncentrationerna av dessa ämnen i alla floder och sjöar inom de Stora Sjöarnas avrinningsområde.

En massbalansmodell för att uppskatta flodvattens koncentration av DOC skapades 1997, men den byggde inte på GIS data. Det finns också en modell som uppskattar hur mycket av inkommande DOC i sjöar som ej når utloppet, dvs som hålls kvar via sediment eller som avgasas från sjön som koldioxid. Modellen kallas Sjö DOC Modellen, LDM (the Lake DOC Model). Dessa modeller tillsammans användes i en tidigare studie för att uppskatta sjökoncentrationer av DOC i 859 sjöar i ett stort avrinningsområde. Mitt exjobb gick ut på att skapa nya massbalansmodeller för att med hjälp av egenskaper hos avrinningsområden uppskatta koncentrationen av löst organiskt kol, DOC, i flodvatten.

Först undersöktes resultat av tidigare studier i Kanada och resten av världen. Detta för att få en bild av vilka egenskaper, utöver våtmarksprocents om användes i modellen från 1997, som kan tänkas förklara flödet av DOC från mark till flod, och därmed koncentrationen av substansen i flodvattnet. Flera andra studier hade också funnit att våtmarker har betydelse, men även lutning, skog, avstånd mellan våtmark och mätpunkt, med flera hade funnits ha betydelse i olika delar av världen. Även egenskaper som kanske kan ses som regionala, såsom geologi och jordtyp, kan ha betydelse.

Då flera kandidater till modellparametrar hittats anhölls om de GIS-lager som krävdes för att kunna beräkna fram dessa för de 20 avrinningsområdena i Dorset, Ontario. Dessa 20 floderna hade bevakats, med avseende på bland annat DOC, under mellan 12-20 år under perioden 1978-1998. Det var också samma områden som användes för att 1997 ta fram ”torv” modellen för flodkoncentration av DOC.

Alla önskvärda parametrar kunde inte erhållas i de lager som var tillgängliga och alla lager var heller ej tillgängliga för området (t ex så täckte inte jordlagret området). 26

parametrar kunde dock beräknas fram från GIS, som t ex genomsnittlig lutning för floden och avrinningsområdet, våtmarksprocent av avrinningsområdet, avstånd mellan våtmark och sjö, flodlängd och procent av area som är skog. Med mätvärden för DOC och beräknade värden för de 26 parametrarna från GIS togs modeller fram för att förklara koncentrationen av DOC i floderna.

Åtta modeller med en till tre parametrar, innehållande totalt sex av de 26 undersökta parametrarna, undersöktes ytterliggare, men bara tre stycken ansågs bra nog. Den första modellen (M1) innehöll endast den genomsnittliga lutningen hos avrinningsområdet, medan nästa modell (M3) innehöll denna parameter och procent våtmark sett till avrinningsområdets area. Den tredje modellen (M8) var ytterliggare en påbyggnad av M3 och innehöll som tredje parameter dräneringsdensitet (vilket är total flodlängd i avrinningsområdet dividerat med avrinningsområdets area). Dessa tre modeller användes för att uppskatta DOC i sjövatten i flodområdet Muskoka. För de 859 sjöarna med en area över fem hektar i området erhölls ett värde för DOC med hjälp av vardera av de tre modellerna tillsammans med Sjö DOC Modellen. Modellen från 1997 användes också med nya våtmarksdata.

I Sjö DOC Modellen finns två parametrar för att beräkna hur mycket av DOC som inte lämnar sjön, så kallade förlustskoefficienter. Denna ena kopplas till inkommande DOC från avrinningsområdet (alltså resultatet från modellerna) respektive från andra sjöar uppströms (deras utflöde av DOC från Sjö DOC Modellen). Dessa koefficienter har inget exakt värde utan för varje modell optimerades deras värden för att få en minskad avvikelse från uppmätta värden av DOC i 237 av de 859 sjöarna, men också separat för de 117 sjöar som var av första ordningen och därmed bara påverkades av den ena förlustkoefficienten (får inget vatten från andra sjöar).

Dessa optimerade värden användes sedan när modellerna uppskattade DOC i sjövattnet. De fyra modellerna förklarade mellan 44- 53 %, beroende på om alla sjöarna användes eller t ex bara sjöar som inte fick vatten från någon annan sjö uppströms (dvs den var den ”högsta” sjön i området). Modellen från 1997 förklarade både lägst 44 % (alla 237) och högst 53 % (endast 90 av sjöarna) av koncentrationerna. Modell M1 var annars sämst på 45 %.

Om resultatet från de olika modellerna sammanställdes till ett medelvärde kunde dock de fyra modellerna tillsammans förklara 50-54 % av koncentrationerna. Detta ger slutsatsen att flera modeller bör användas för att få en bättre uppskattning och mindre spridning av resultatet. Med flera modeller kan man också erhålla en spännvidd av koncentrationer och inte ett enda värde.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANC	Acid neutralization capacity
$CFCl_3$	CFC-11, greenhouse gas
CF_2Cl_2	CFC-12, greenhouse gas
CH_4	Methane, greenhouse gas
CO_2	Carbon dioxide, greenhouse gas
DEM	Digital Elevation Model
DIC	Dissolved inorganic carbon
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
DON	Dissolved organic nitrogen
ESRI	Environmental Systems research Institute (development of GIS)
FA	Fulvic acids
FRI	Forest Resource Inventory
GCM	Global Circulation Model
GIS	Geographic Information System
HA	Humic acids
HS	Humic substances
hw	Headwater lake
ILDB	Inland Lake Database
INCA-C	Integrated Catchments model for Carbon
LDM	the Lake DOC Model
LiDAR	Light detection and ranging
LMW	Light molecular weight
MD	Muskoka District
MNR	Ontario Ministry of Natural Resources (sometimes also OMNR)
MOE	Ontario Ministry of the Environment
MRW	the Muskoka River Watershed
MWCI	Muskoka Watershed Council Information
N_2O	Nitrous oxide, greenhouse gas
NAD_83	North America Datum 1983 (used in ArcGIS)
NRVIS	Natural Resource and Values Information System
OBM	Ontario Basic Mapping (also used for one of the wetland layer)
OLS	(Ordinary) Least Square

POC	Particulate organic carbon
Q	Runoff
RAT	Rapid Assessment Technique
SA	Sensitivity analysis
SOM	Soil organic matter
SPSS	Statistical software
S-PLUS	Statistical software
THM	trihalomethanes (carcinogenic substance)
TOC	Total organic carbon
UTM	Universal Transverse Mercator
UV	Ultra-violet light
VB(A)	Visual Basic (for Applications)
VIF	Variance inflation factors

Abbreviations in equations:

abbreviation	meaning	used
[DOC]	Mean annual concentration	
DOC_{est}	Estimated valued of DOC	For MRW
DOC_{in}	Inflow of DOC to a lake	
DOC_{m}	Measured/observed value of DOC	For MRW
DOC_{out}	Outflow of DOC from a lake	
ε	Random/unexplained error	
L	Load	
L_o	Loss via outflow	
R	Retention	
R_{doc}	Net retention of DOC	
q_s	Areal runoff	
v	Net loss coefficient	
v_l	Loss coefficient for the catchment to the lake	Lake DOC Model
v_u	Loss coefficient for upstream lakes	Lake DOC Model
z	Mean depth	

Abbreviations and meaning of parameters from GIS:

Parameters x_i	Units	Meaning of parameter	*
area_peri	m	Catchment area/catchment perimeter	4
avslope	° **	Average slope in subcatchments	1
catarea	m^2	Catchment area	3
catperi	m	Catchment perimeter	2
distlakeOBMav	m	Average straight distance between lake and OBM wetland	10
distlakeOBMmax	m	Max straight distance between lake and OBM wetland	11
distlakeOBMmin	m	Min straight distance between lake and OBM wetland	12
distlakeRATav	m	Average straight distance between lake and RAT wetland	13
distlakeRATmax	m	Max straight distance between lake and RAT wetland	14
distlakeRATmin	m	Min straight distance between lake and RAT wetland	15
drainden	m^{-1}	Drainage density = stream length/catchment area	24
lakearea_cat	-	Lake area/catchment area	22
perFOR	%	Percent of forest cover on catchment	7
perFOROBM	%	Percent forest cover on OBM wetlands	8
perFORRAT	%	Percent forest cover on RAT wetlands	9
perOBM	%	Percentage of OBMwetlands (from NRVIS)	5
perOBM2	%	Percentage of OBMwetlands + ponds from Ducks unlimited	20
perOBM3	%	Percentage of OBMwetlands + ponds from Bata Library	21
perRAT	%	Percentage of RATwetlands (from Ducks Unlimited	6
perRAT2	%	Percentage of RATwetlands + ponds from Ducks unlimited	18
perRAT3	%	Percentage of RATwetlands + ponds from Bata Library	19
perRoad	%	Percentage of road on catchment area	23
spond	%	Small lakes/ponds from Ducks unlimited	16
Strslope	°	Average stream slope	25
Strslope_len	°/m	Average stream slope / stream length	26
wpond	%	Small lakes/ponds from Bata Library	17

*Numbers used in correlation matrix, see Appendix D

** Degrees = °

1 INTRODUCTION

Dissolved organic carbon (DOC) is important because it affects the flux, concentration and toxicity of metals and nutrients in aquatic systems. This results, in part, from the fact that nutrients and metals like mercury and lead form complexes with DOC that result in the co-export of these substances from soil to streams and lakes. DOC is also a part of the global carbon cycle, linked with the concentration of CO_2 in the atmosphere and thus also the climate changes that follows. The use of chlorine in water treatment can result in reactions between the chlorine and DOC that will produce a carcinogenic substance.

1.1 AIM

In this project I aim to update a black box, mass balance model of the flux of DOC based on catchment properties. I will investigate the flux of DOC in a tertiary watershed in the Great Lakes Basin in Canada, by using the updated model(s) on each individual catchment. I will evaluate whether the parameter(s) in the new model(s) explain as much or more of the flux of dissolved organic carbon within the catchments. These results will hopefully add to the knowledge of what factors are controlling the flux of DOC within a catchment and how changes in the climate could affect these relationships.

This project is a part of a bigger project that aims to explain not only flux of DOC but also other fluxes and concentrations of contaminants and nutrients in the aquatic environment. The bigger project also aims at gaining more knowledge into the consequences of changes due to acidification (or recovery from acidification), climate change and other processes on elemental fluxes. One goal of the major project is to produce a model that can explain the differences in flux of DOC between catchments within the boreal forest and to be able to use this on the whole of the Great Lakes Basin area. For this to be possible the input to the model – i.e. the parameters of the model – needs to be available at a larger scale. One type of data that is available for large parts of the world is GIS-based land cover.

The goal of this work is to find at least one parameter that can be obtained using GIS that explains a significant portion of the DOC flux. The hope is that GIS information now has good enough accuracy such that a new model will be as good or better then previous models that were based on land-use data from a combination of air photos and field work. The main aim is to investigate if GIS data is at present good enough to use as model input. My goal and aim will be reached when parameters relevant to the flux of DOC can be found in GIS data and used in models to estimate the DOC concentrations in the large tertiary watershed. The main aim will be reached only if the/those model(s) with GIS data as input are as good or better at estimating DOC than older models based on none GIS data.

2 BACKGROUND

Dissolved organic carbon, DOC, is carbon from organic sources that is dissolved in water. The flux of DOC is an important part of the global carbon cycle and contributes to the production of atmospheric CO_2. The increasing amount of CO_2 in the atmosphere is the major cause of climate change, a fact now recognized by the majority of the scientific community (Issar, 2004). Other components of the carbon cycle therefore become important, as they may act as sources or sinks for atmospheric CO_2 and may themselves be affected as the climate changes.

DOC present in waters in soil, streams and lakes forms complexes with metals and other contaminants as well as with nutrients. The flux of DOC can therefore explain parts of the fluxes of other substances as well as their toxicity/bioavailability.

There have been several models for the flux of DOC proposed previously. These utilize catchments characteristics, chemistry in waters and soil and the connection between DOC in waters and the colour of the water. This present work focuses on developing a black box, mass balance model suitable for Precambrian catchments in Ontario, Canada. Data measured in the Dorset study during 20 years will be used and the model developed then applied to the whole of a large tertiary catchment, the Muskoka River Watershed. The input will come from available GIS information on catchment characteristics, with the intent that this will make the model applicable over a larger area as measurements or other field data will not be necessary.

2.1 THE GLOBAL CARBON CYCLE

Carbon is one of the basic elements in nature and it is one of the main constituents of organisms. The global carbon cycle (for a simple picture see Figure 1) involves the gases in the atmosphere, the carbon in animals and vegetation in the biosphere but it also involves the carbon in the soil which is transported to the streams, lakes and sea.

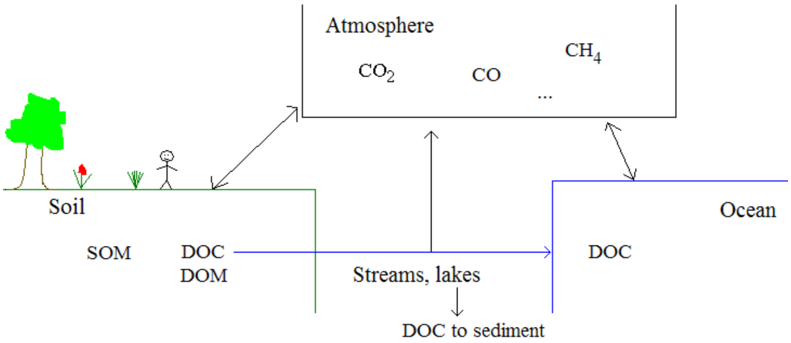


Figure 1. Parts of the global cycle for carbon. DOM stands for dissolved organic matter, SOM for soil organic matter and DOC is dissolved organic carbon.

2.1.1 Soil organic matter

In the biosphere, primary production forms organic matter. As organisms die, leaves fall, roots and animals die, the organic matter partly or entirely end up in the soil or in new organisms. Microbes, like bacteria and fungus, break down the organic matter, which leads to the formation of more stable organic matter that is not so easily decomposed. The decomposition process also releases inorganic nutrients that become available to soil organisms and plants for primary production of new matter. (Gustafsson *et al.,* 2005)

Soil organic matter (SOM) may be divided into different kind of groups, for example into an active and a passive group, referring to their “status” towards decomposition. Often the more stable high-molecule-weight-bi-products of decomposition making up the latter group is referred to as humic substances (HS) or humic matter. Humic matter is present in soil, water of streams, lakes and oceans and in their foams and sediment, in every ecosystem on earth. It is also a big part of depositions of for example peat, oil shale, and fossil fuels. (Tan, 2003)

Humic substances have a major effect on the properties of soils. Different soil types contain different amounts and composition of humic matter (Tan, 2003). SOM is available for plants mainly in dissolved form, as dissolved organic matter (DOM).

2.1.2 Dissolved carbon

The microbial degradation of SOM, then followed by desorption of organic substances from soil, leaching of organic substances from fresh litter are together thought to be the main processes for release of DOM (Michalzik *et al.,* 2001).

Often DOM is used interchangeably with DOC, dissolved organic carbon (O'Connor, 2007), or TOC, total organic carbon, despite the fact that they also contain other substances (like phosphorus and nitrogen). TOC is actually divided into DOC and POC (Particulate organic carbon), but DOC often constitutes more than 95 % of the TOC (Futter, 2007). By definition DOC is the organic carbon that can pass through a 0.45 µm filter (Creed *et al.,* 2003; Futter, 2007; O'Connor, 2007), excluding most bacteria, plankton and particulate matter (like POC).

Humic substances accounts for between 40-60 % of DOC in lake water (Creed *et al.,* 2003; O'Connor, 2007). DOC contains organic compounds that range from low weighing simple amino acids to the higher weight fulvic and humic acids (Molot and Dillon, 1997a; Creed *et al.,* 2003). DOC's wide variety of compounds (Moore 2003) has differing physical and chemical properties, but is often treated as an average composition (Dillon and Molot, 1997b; Michalzik *et al.,* 2001; Neff and Asner, 2001). It can be divided into different fractions, for example after solubility in acids and alkaline agents, called fulvic and humic acids and humin (Tan, 2003; Gustafsson *et al.,* 2005; Futter, 2007). Hydrophobic and hydrophilic acids are also mentioned as the main fractions of DOC. Most of the different divisions of DOC have a specific purpose. For example: the hydrophobic fraction of DOC contains almost all the aromatic components of DOM, while the hydrophilic is mineralized faster, but sorbed less strongly relative to the hydrophilic fraction. (Dillon and Molot, 1997b; Michalzik *et al.,* 2001; Neff and Asner, 2001) Also the Humic acid (HA) fraction is more resistant to change than the Fulvic acid (FA) fraction, which therefore has a lower concentration in streams and lakes than in soil and wetlands, as it is broken down. (Futter, 2007)

2.1.3 Dissolved organic carbon in streams and lakes

The basic types of DOC in streams and lakes are (this is another division of DOC, this time depending on the source) (Lindsjö, 2005):

- ❑ Allochtonus – terrestrial production of DOC. This is the largest group (Molot and Dillon, 1997b; Tan, 2003; Futter, 2007). Comes via groundwater/subsurface flows or surface runoff to the stream and/or lake.
- ❑ Autochthonous – produced within the system. Primary production and breakdown of algae (Tan, 2003; Futter, 2007). Is decomposed fairly quickly and makes up a small part of the surface water DOC (Dillon and Molot, 1997a; O´Connor, 2007).
- ❑ Anthropogenic – human sources (and sinks): Industry, agriculture, domestic sources. Less is known of these sources and the composition of the DOC from them (Tan, 2003).
- ❑ Atmospheric DOC – not a big source. Can for example come from deposition of biological material (e.g. pollen) to streams and lakes.

As the DOC has left the soil and entered the streams, as subsurface- or groundwater flow, it is affected by factors, like radiation. DOC is affected as single components and as a whole, as concentration of DOC and of a changing distribution of its different fractions. In the stream and lake production, biotic processes and adsorption of DOC occur (O'Connor, 2007). Solar radiation may decompose DOC and different types/parts of DOC may be less or more susceptible to light and other factors altering the fractional parts of DOC as well as how one might relate colour to DOC concentration.

2.1.4 Effects of the DOC flux on soil and water ecosystems

DOC is present in all ecosystems (Neff and Asner, 2001), and for the lake-stream-land ecosystem and the biogeochemistry in terrestrial and aquatic systems it is of great importance. One important part is the cycling of DOC within soils, which is important for soil formation, distribution of substances and stabilization of soil carbon as a whole (Neff and Asner, 2001; Futter, 2007).

It is of great importance to know which factors affect the flux and composition of DOC to be able to see how changes due to for example changes in climate and acidification may affect DOC. Runoff is highly correlated with DOC flux from soils (Neff and Asner, 2001) so changes in runoff may affect the amount and concentrations of DOC in soil, streams and lakes. Runoff depends on temperature and precipitation as well as many other climate factors. DOM contains carbon as well as essential elements, like nitrogen, sulphur and phosphorus and is therefore an important source of energy (Findlay *et al.,* 2001; Futter, 2007) for aquatic life downstream and a decrease in the inflow can effect the capacity for the ecosystems to support primary production (Creed *et al.,* 2003). One more part affecting primary production can be changes in some properties of the physical environment that are also affected. For example it may alter the penetration of UV-B light, which is especially affected by the coloured fractions of DOC. More UV-B light can be damaging for the organisms living in these waters and changes the primary production and heat storage in the waters (Dillon and Molot, 1997b; Schiff *et al.,* 1998; Creed *et al.,* 2003; Hudson *et al.,* 2003; Mulholland, 2003; Dillon and Molot, 2005; Futter, 2007).

As some trace metals, organics and nutrients bind to DOC the export of DOC can also affect the export of these substances from the soil. The toxicity of these substances within the system is also affected by the concentration of DOC since complex species may be less harmful (Schiff *et al.,* 1990; Boyer *et al.,* 1996; Dillon and Molot, 1997b; Schiff *et al.,* 1998; Hudson *et al.,* 2003; Mulholland, 2003; Futter *et al.,* 2007; O'Connor, 2007). It may also be a health problem as many water treatment plants use chlorine (Tan, 2003) as part of the treatment, and chlorine reacts with DOC and forms carcinogenic trihalomethanes (THM) (Futter *et al.,* 2007).

Changes in the rate of DOC loading/flux into and out of a system like a lake can influence many of the water chemistry parameters (Futter, 2007; O'Connor, 2007). The acid-base balance of the aquatic system could for example be is shifted. Organic acid anions that are a part of DOC can account for up to 20 % of the total acid neutralization (buffering) capacity (ANC) of a lake (Schiff *et al.,* 1990). DOC has a number of weak acid functional groups – carboxylic acid, anolic hydrogen, penholic OH for example – and can also buffer inputs of strong inorganic acids (weak acids is matched by strong bases) (Futter, 2007).

2.1.5 Dissolved inorganic carbon

There is also dissolved inorganic carbon (DIC), which is connected to DOC as DOC mineralizes to DIC. DIC also acts as a main acid buffer, affecting ANC, in for examples forested lake watersheds in Canada. The study of Avarena *et al.* (1992) investigated the production and cycling of DIC by measuring stream and lake DIC in the study area, for the most part Harp Lake with its six subcatchments (also used in this study, see section 3.1). Since DIC is largely of importance in sedimentary areas (i.e. carbonate bedrock), it is not of great importance in this study.

2.1.6 Climate change

A drier climate can lead to less wetland proportions and thereby less DOC but also more fluctuations in the DOC levels (Schiff *et al.,* 1998). Warmer and drier climate will probably reduce the flux of DOC, diminish the area of wetland as a result of more evaporation and less precipitation (Mulholland, 2003).

With rising temperatures the hydrological cycle gains new energy and reaches a new pace. Vegetation regimes might be altered, plants and animal having to adapt, move or become extinct. Attempting to estimate the effect of the increase in greenhouse gases leads to the use of global models. These are called GCM´s, Global Circulation Models, Global dynamics models (Schnoor, 1996) or General Climatologically Models (Issar, 2004) and use the fact that different greenhouses gases like methane (CH_4), nitrous oxide (N_2O), CFC-11 ($CFCl_3$) and CFC-12 (CF_2Cl_2) can be transferred to CO_2-eq (measuring their effect relative to the effect of CO_2).

Temperature and changes in the hydrology affect the vegetation, evaporation, precipitation, and soil moisture directly and indirectly. Theses changes affect the flux of substances like DOC in the ecosystems. Knowing how each source of DOC react to local variations in the weather and how they are contributing to the flux of DOC gives scientists a chance to estimate the effect the changes will have on DOC and also CO_2. The load of DOC to streams and lakes is affected, but the fate of DOC in lakes will also change as water residence times are altered and the ratio between the paths in which carbon is divided can change. This in itself can add or subtract to the concentration of CO_2 in the atmosphere as the catchment as a whole is a sink or source for carbon.

2.2 FACTORS AFFECTING EXPORT OF DOC FROM SOIL

Factors affecting the export of DOC from soil can be both regional and local. The regional factors are for example climate, as it affects a large region equally. The effects of these factors are best studied on areas that are similar but situated far apart. If the studied areas are under different climate factors, but are similar in catchment size, lake size, and so on, the effect of only the regional factors can be investigated. For example the study conducted by Fröberg *et al.* (2006) looked at DOC in different horizons in three boreal forest locations in Sweden. The three sites had many characteristics in common but were down a climate gradient, in that they where from south to north and had different average temperatures.

The research of this project is focused on local factors as the areas being studied (both for building the model and for using it) are in close proximity to each other. To understand how the flux of DOC is controlled by the hydrogeology (local factors) in a catchment, different factors must be considered. Looking at different smaller catchments within the same area can give a picture of which factors create a high flux or a low flux of DOC (Dillon and Molot, 1997b; Creed *et al.,* 2003). Different studies have focused on different factors and some factors have been found that seem to affect or not affect the DOC flux.

No matter which type of factor the research is focused on, modeling the DOC fluxes and concentrations of lakes and other surfaces water systems may give new knowledge of what factors affect the flux, to what extend they do so and may also be a tool to estimate the effects of changes in a system due to acidification, climate changes and so on. (O’Connor, 2007)

2.2.1 Wetlands

The principle source of DOC/DOM in boreal ecosystems is the catchment and in particular the wetlands (Molot and Dillon, 1997b; Creed *et al.,* 2003; Dillon and Molot, 2005; Wu *et al.,* 2005; O’Connor, 2007). “Wetlands are the principal sources of dissolved organic carbon (DOC) to streams, rivers and ultimately lakes in forested ecosystems” (Creed *et al.,* 2003). The older mass balance model (Dillon and Molot; 1997b) had only peat (wetland) percentage as a factor.

Wetland areas exist in all regions of the world, from the tundra to the tropics. Peat or wetland areas are areas that experience poor drainage and where anaerobic decomposition therefore prevails (Tan, 2003). The saturated state of the soil in a wetland leads to an accumulation of carbon (Schiff *et al.,* 1990), which can later leak from the area with subsurface flow of water, to streams and lakes (Dillon and Molot, 1997b). DOC also percolates down in unsaturated soil but most of it stays in the soil profile because of adsorption in the mineral horizons (O’Connor, 2007).

The anaerobic state can affect the composition of DOC. There are different definitions of organic soils (which are peat, muck and so on), many of which are given by Tan (2003, chapter 2). Wetlands themselves are also divided into different types depending on the different factors forming them. In Ontario the main parts of wetlands are divided into (O’Connor, 2007):

- Bogs: Acidic, rich in peat and plant residue. Water mostly comes from precipitation.
- Fens: Alkaline, accumulate peat deposits. Marsh like vegetation. Fed by groundwater.
- Marshes: usually saturated or seasonally flooded with other water than rainfall. Grasses and herbaceous plants.
- Swamps: low topography and at least seasonally flooded. More wooded plants than marshes

Creed *et al.* (2003) investigated wetlands hidden beneath the forest floor/canopy, called cryptic wetland, and their effect on the DOC flux. According to this study the presence of wetlands could explain about 90 % of the natural variation of average annual DOC export in the investigated catchments, which was the Turkey Lakes Watershed in central Ontario. This watershed contains only a few wetlands, but more of the area could be seen as cryptic wetland. One main conclusion of the study was that for DOC exports models both the cryptic and non-cryptic wetlands should be a part. The cryptic wetlands can be found with different methods, manually or with GIS (Geographic Information Systems) using the topography given by DEM (Digital Elevation Model) (Creed *et al.,* 2003). These DEM have to have a high accuracy though, both vertically and horizontally (which is not commonly available).

2.2.2 Other local factors

The goal of this project is to try and find other local factors, apart from wetlands, available from GIS data that may explain some significant part of the flux of DOC. Many factors may affect the flux of DOC, positively or negatively, but not all of these can be obtained from GIS data or other data available for large areas. This makes them non-useful in the attempt to model DOC flux on a larger scale.

Many studies have been done, some also involving modelling to see what can be used to predict DOC flux. For example Mulholland (2003) looked at parameters like: channel slope, watershed/catchment slope, mean lake depth, lake area, water residence time, drainage area/lake perimeter ratio and conifer abundance.

In 1997 a model was obtained with data from the Dorset study (see section 3.1). The study of Dillon and Molot (1997b) looking at a number of factors that could affect DOC exports, meteorological, hydrological, and physiographic aspects as well as bedrock geology. Within these they examined variables like: catchment area, average catchment grade (%), stream length, % area as pond, exposed rock, mean annual air temperature, relative humidity and many others. What they could really relate to DOC export, with their 0.15 significance level, was peatland percentage of the catchment area. This accounted for about 78 % of the variance in a stepwise regression model: [DOC] = 2360 + 261 · (% peat). This model will be used in this project to compare the result of the models that will be developed.

Lindsjö (2005) used map information to model DOC in Sweden and looked into for example the following parameters in his study: Catchment area, stream length, drainage density ([total length of streams within a catchment]/[total area of the catchment], gives a measure of the average lateral flow path length through soil to the stream network.), sinuosity ([stream length]/[shortest distance between two sampling sites], is a measure of the streams crookedness, if it meanders or is straight), slope, elevation, arable field, forest, forest clear cut, open land, pasture, water, wetland (forest, impassable, open, total, within ten meter of stream), soil types, bedrock, age of forest stand, average height of forest, volume of different species of forest, lake length.

The work of Bishop *et al.* (1994) looked into the riparian zones (soils near stream) as sources of aquatic DOC. The results showed that the zones delivered DOC, but how much the different types of riparian zone soils exported was harder to quantify. The work of Findlay *et al.* (2001) in New Zeeland also looked into riparian zones as a source of DOC, as well as the effects of land use. The result was that land use affects the DOC as do the riparian vegetation, the latter since shadowing from vegetation can affect the amount of solar radiation that reaches the surface of the stream and thereby also affecting the decomposition rate of DOC. It was also found that the DOC level was mostly dependent on the land use about 50 years ago (which can have something to do with the findings that DOC from for example wetlands is quite resent, about 40-45 years old, see section 2.2.5.). How the DOC reacted to different levels in solar radiation was also dependent of the land use, probably a sign of different DOC compositions.

Vidon and Hill (2004) studied the landscape control over hydrology in riparian zones in southern Ontario, in some agricultural catchments. What was found was that there was somewhat of a threshold when it comes to slope of the riparian area. Topography affects the flow path, but also stratigraphy and hydraulic properties of the soil have an influence. One important feature is the presence or absence of a confining layer at some, not too deep, depth in the riparian zone. These together affect the hydrologic connection with upland area, which in turn affect the direction of flow.

The study of Michalzik *et al.* (2001) found that 46-65% of the annual flux of DOC and DON (dissolved organic nitrogen) could be explained by fluxes of DOC and DON in throughfall in their study.

The study of Moore (2003) found that, forested fires affect the level of DOC, in the soil and in the precipitation. Sorption of DOC depends of the composition. Like Schiff *et al.* (1998) and Futter *et al.* (2007) this study also mentions that the presence of open ponds within the wetland area decreases the DOC flux from the wetland as a whole.

2.2.3 Correlation between local factors

Properties within a catchment are to some extent correlated. This can be connected to the fact that they are formed under the same regional factors (making regional and local factors correlated too, which is one reason why areas chosen for a study in regional factors should be as similar as possible in as many aspects as possible). For example slope and wetlands are correlated as wetlands are formed in low flat areas in the topography. This was found in the study of Dillon *et al.* (1991) that also mentioned that the typical forests types for well-drained soils are deciduous or mixed forests and for poorly drained soils mixed or coniferous forest. This means that the types of forest are also correlated to wetlands and slope. Many other connections, coming from direct and indirect effects that one factor have on another, can be found.

2.2.4 The climate change effect on catchment properties

A change in climate means a change in the regional factor that has been affecting an area. The change in the climate will also affect local factors connected to the regional climate. As changes affect the catchment properties that have an influence on the export of DOC, the latter will also be affected. More and more studies look at estimating the effect of climate change on different aspects of ecosystems and so also the export of DOC. For example the study of Magnuson *et al.* (1997) which looked at the potential effect of climate change on the Precambrian Shield. The authors mention in the article that decreases in DOC input should be expected from drier catchments, as temperature increases and precipitation decreases. Vegetation regimes will shift northwards as will fish communities for example. This will affect the whole systems in complex ways. The study looked at different model simulated scenarios and the effects on different systems during recent droughts. One important thing mentioned is the increase in lake water retention times and decrease in lake area and volume. Some lakes might even disappear as the water input decrease from upstream lakes, precipitation decline, or though a hydrologically disconnection from groundwater inputs.

Schiff *et al.* (1998) studied wetlands in the Precambrian Shield aiming at gaining a better foundation when one wishes to estimate the effect of climate change. It was found that a drier climate would give less DOC for example, due to lower water tables which could give less wetlands or disconnected wetlands. The effect of the lower DOC would be clearer waters, with less cold water, changing the depth of the thermocline and also affecting the level of solar radiation reaching different depths.

The study of Michalzik *et al.* (2001) mentioned effects on pH as DOC fluxes seemed related to this property. At higher pH values there might actually be a more favorable environment for the decomposers in the soil leading to more DOC being released from SOM, but an increased deprotonation of functional groups would also give a higher solubility for DOC. The effect is independent of which mechanism is responsible, a higher DOC concentration at higher pH. As climate change can affect pH this is another way in which it can affect the DOC in soil and water.

The risk of lakes and ponds losing some of their biodiversity as DOM decreases and more UV-B light reaches the lake water at different depths was investigated by Molot *et al.* (2004). It was pointed out that other factors besides colour also are affected. For example a climate change with higher temperatures leads to a higher evapotranspiration rate that can lower the water surface. Even with just the DOM and the colour, the relationship between them is not stable but varies to such a high degree that a 50 % change in DOM might mean that the coloured part decreases 40, or 60 %.

The main result is that the effects of climate change on DOC are complex and different systems may respond differently to a similar change.

2.2.5 The age of DOC in lakes

The DOC in one location will have an average age of all DOC that has reached that location from all sources located above that point in the catchment. Even with knowledge of the importance of DOC little is known of the production and turnover of DOC within natural watersheds. The studies of Schiff *et al.* (1990 and 1997) were looking into the turnover times of DOC (^{14}C) and also the possible sources within the system for DOC and the fractions of DOC from these (^{13}C). The studies were looking into the age (^{14}C) and source (^{13}C) of DOC in a number of catchments in the Precambrian area in the Muskoka District, Ontario, by measuring the ^{14}C and ^{13}C (source of C different for C3 and C4 plants, which differ in the way the plants first assimilate CO_2 from the atmosphere (www, SERC, 2008)) of the carbon from different parts of the system: the groundwater, the streams, soils, sediments and the lakes. DOC coming from groundwater was older and had a lower concentration, while water from shallow subsurface flow was younger. Schiff *et al.* (1997) concluded that about 50 % of the DOC was less than 40 years old (45 years according to the study of Schiff *et al.* (1998)). The age also varied within one catchment area over space and time, mainly with seasons and storm events. The different sources of DOC also gave different seasonal patterns as wetland dominated catchments had less seasonal differences. The relative proportion of DOC from wetland compared to upland area also changed seasonally. Many storm events in catchments where the age of the DOC was old meant that the DOC flux increased as riparian zone close to the streams where flushed. If the DOC was already young the flushing of the storm event was not as important. The sources in the catchment were named as; wetlands, riparian zones near the streams (more at high flows), groundwater (small), beaver ponds, and in-stream production. (Schiff *et al.,* 1997)

It seems that most of the DOC that reaches the stream from the wetland is quite resent (Dillon and Molot, 1997b). This suggests that the flow paths are close to the surface of the soil. The peat that is buried at larger depths is also resistant to mineralization and mobilization. The resent age is also consistent with the retention of DOC by mineral soils (Dillon and Molot, 1997b).

2.3 FACTORS AFFECTING THE FATE OF DOC IN STREAMS AND LAKES

The DOC can be changed by light, be photo bleached, chemically altered to DIC (dissolved inorganic carbon) to mention some of the processes that affect the DOC. By both abiotic (adsorption, flocculation (Molot and Dillon, 1997a)) and biotic (uptake by microorganism, which leads to respiration of CO_2) processes the DOC is thus removed from streams and lakes. From lakes the remaining DOC then leaves with the out flowing water, which tends to have a lower concentration than the incoming water due to the losses in the lake.

2.3.1 Colour and photo-oxidations

The allochtonus (terrestrial) carbon has a higher photo-oxidation rate than the autochthonus (Genning *et al.,* 2001). The organic carbon that has been changed by light into LMW (low molecule weight substances) which are more available for biologic uptake (Molot and Dillon, 1997a), see Figure 2.

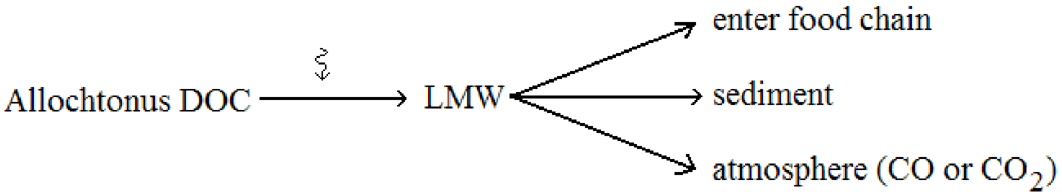


Figure 2. The fate of allochtonus DOC as it is in a light exposed environment. LMW stands for low molecular weight. (Genning *et al.,* 2001).

TOC (total organic carbon) in Ontario lakes is lost to sediments or degraded (for example via UV radiation) (Genning *et al.,* 2001) and as a result of the latter lost to the atmosphere as mainly CO_2, but also CH_4. The partitioning between sedimentation and losses to the atmosphere depends on the acidity/alkalinity of the lake (Molot and Dillon, 1996; Genning *et al.,* 2001). Sedimentation and evasion both follow after photo degradation (Wu *et al.,* 2005)

The study of Jonsson *et al.* (2007) in Sweden showed that sedimentation explained 3 % of the loss of carbon, evasion 45 % and 50 % was exported to the sea (where evasion of CO_2 to the atmosphere can continue). Most of the accumulation in the system was due to build up of tree mass, whereas clear cut areas were sources of carbon.

The study of Molot and Dillon (1997a) that looked into the photolytic and non-photolytic decomposition of DOC showed that the amount of DOC that did not leave the lake, but instead evaded to the atmosphere or sedimented, was 38-70 % of the DOC load in the seven lakes studied between the years 1980-1992. The study also showed that lake DOC was not as affected by light as stream DOC.

The study of Köhler *et al.* (2002) was focused on light and microbial activity decomposing TOC in water from soil, lakes and streams. Much of the TOC in water samples that were exposed to light treatment ended up as CO_2. During this the pH and the alkalinity of the water increased, the latter contributing to the ANC that therefore was strongly related to the amount of TOC. The remaining TOC had a lower average molecular weight, so the composition of the TOC was changed, and how it was altered depended on the source of the water/TOC. The study made by Genning *et al.* (2001) also found an alkalinity increase as TOC decreased. They found that more carbon was going to the atmosphere compared to being sedimented. The sedimentation goes down and atmosphere evasion goes up when the lakes is acidic (Wu *et al.,* 2005).

It seems that oxidation also can occur with the help of the photo-oxidants, in the form of hydroxyl radicals, OH•. In the study of Molot et al. (2005; Wu *et al.,* 2005) the fate of DOC in the pH interval 4-9 was studied. The importance of OH•, decreased as the pH increased until it was negligible.

In sediments carbon is stored as POC, particulate organic carbon (Molot *et al.,* 2005). Mostly high molecular weight DOC is sedimented, and low molecule weight (LMW) DOC can later be released back into the water (Molot and Dillon, 1997b).

3 STUDY AREAS

Two areas were used in this study. The first area contained seven lakes in Dorset, Ontario, where a total of 20 streams have been measured for DOC during 12-20 years in the period 1978-1998 and most are still being monitored. Some of the lakes in this area lie within the second area, the Muskoka River Watershed, where 859 lakes have been delineated (their catchments computed in GIS) for an earlier Master Thesis work (O´Connor, 2007).

The first area was used to derive the mass balance models and those models deemed the best were used to estimate the DOC in the lakes of the Muskoka River Watershed, together with a Lake DOC Model that takes the stream DOC concentration and transfer it into lake DOC concentration (see section 4.1.2) and made it possible to connect the whole watershed.

3.1 LONG-TERM STUDY IN DORSET

This area consists of seven lakes (see Figure 3 below and Figure App 1-7 in Appendix A) and the 20 subcatchments derived from where streams DOC concentrations have been and is being measured. The streams are all a part of the Dorset long-term study and have a lot of data dating from 1978 (or some years later) and forward (most had 20 years of data, but one stream had only twelve years of data, for the number of years for each subcatchment see Table App-1 in Appendix A). More on how the data were derived for the earlier as well as for this present study are available in Molot and Dillon (1997a; 1997b; Dillon *et al.* 1991). Data from the report of Dillon and Molot (1997b) is also shown in Table 1 below and in Table App-1 in Appendix A.

The focus of the Dorset study was to learn more about the impacts of long-range atmospheric transport of for example substances that are a part of anthropogenic acidification, climate change as well as the effect of cottage development on the quality of water (Dillon *et al.,* 2003). This area was also used to derive the original mass balance model (data used to gain this model, as well as subcatchment area are available in Table App-1, in Appendix A), which contains the wetland percentage of the catchment as a way of explaining the flux of DOC. (Dillon *et al.,* 1991; Molot and Dillon, 1996; Molot and Dillon, 1997a; Molot and Dillon, 1997b; Dillon *et al.,* 2003; Dillon and Molot, 2005; Wu *et al.,* 2005)

All the 20 subcatchments are located in close proximity to each other in central Ontario, Canada. They lie with in the county of Haliburton or the Muskoka District (MD) (Dillon and Molot, 1997a) and are a part of the Precambrian Shield. They are forested (Wu *et al.,* 2005) and contain 0-25% wetlands. The streams are of first or second order with a mean runoff of 0.5 m yr^{-1}. The seven lakes are oligo- to mesotrophic and six of them are headwater lakes (meaning that the lake receives no water from any other lake, their "lake order" is 1). The exception to this is Red Chalk Lake, which gets water from Blue Chalk Lake (also in the study) (Molot and Dillon, 1997a). As mentioned some of the lakes/catchments lies within the larger watershed and in Figure 3 the outline of the Muskoka River Watershed (see section 3.2) is also seen and as can been noticed in the zoomed part of the picture, where lakes are also shown, four lakes are actually outside of the MRW and three inside. (Crosson catchment area seems to have more then one lake, but the south one is the actual lake, the other is a small lake, seen as a pond in this study.) (Dillon *et al.,* 1991; Molot and Dillon, 1996; Dillon and Molot, 2005).

The soil cover is generally less than one m thick, and in most locations the soil cover is less than ten m. The bedrock below is Precambrian metamorphic plutonic and volcanic silicate (Molot and Dillon, 1997a; Dillon and Molot, 2005; Wu *et al.,* 2005). The most dominant soil types are brunisolic and podzolic, but due to the extent of wetland, organic soils (peat) are also common. (Also Dillon *et al.,* 1991; Molot and Dillon, 1996; Dillon and Molot, 2005; Wu *et al.,* 2005)

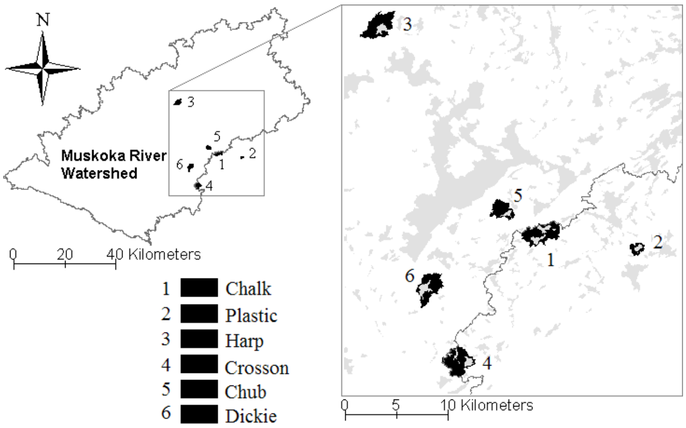


Figure 3. The catchments for the seven lakes used to attain the models are here numbered (number one is two areas, Blue chalk (the northern) and Red chalk lake) and the outlined area is the Muskoka River Watershed (see section 3.2). (Picture made in ArcGIS 9.1 and Microsoft ® Paint 1.5.)

Table 1. Data over the seven lakes, where A_o is lake surface area, A_d catchment area, not including A_o, and $\bar{z}$ mean depth, and DOC is the whole lakes concentrations and mean Secchi depth comes from measurements made during the years 1977-1989 (in Crosson starting in the year 1980 and Plastic 1979).

Lake	Sub catch-ments	A_o [ha]	A_d [ha]	$\bar{z}$ [m]	DOC [mg L^{-1}]	Colour	Colour/ DOC	Secchi depth [m]
Blue Chalk	1	52.35	105.9	8.5	1.8	6	3.3	6.8
Chub	2	34.41	271.8	8.9	4.8	46.5	9.9	3.3
Crosson	1	56.74	521.8	9.2	4.1	35.7	8.5	3.6
Dickie	5	93.60	406.4	5.0	5.0	45.8	9.2	2.8
Harp	6	71.38	470.7	13.3	3.9	21.1	5.7	3.8
Plastic	1 (6*)	32.14	95.5	7.9	2.3	7.9	3.6	6.8
Red Chalk	4	57.13	532.4	14.2	2.5	11.7	4.7	6.3
Sum	20							

Sources: Dillon and Molot, 1997a; Molot and Dillon, 1997b

* Of the six streams going to Plastic Lake and its following six subcatchments there was data only from one, PC1.

3.2 MUSKOKA RIVER WATERSHED

Once the new, updated mass-balance model has been attained it will be used on the entire watershed called the Muskoka River Watershed (MRW), which is a part of the Great Lakes Basin in Canada. It is a tertiary watershed located in south-central Ontario (Figure 3 and Figure 4) and centered (the placement of the centroid of the Muskoka River Watershed, attained from ArcGIS 9.1) at –79.2° longitude and 45.3° latitude, about 180 km, almost straight north of Toronto. It is a quite large catchment as it covers over 5 000 km^2 and the rivers themselves stretch over 210 km and drop 345 m before reaching the final outlet which is Georgian Bay (O'Connor, 2007; www, MWCI, 2007). The 859 lakes in the MRW are divided into three drainage systems, the North and South branches of the Muskoka River and the Lower Muskoka sub-watershed (O'Connor, 2007; www, MWCI, 2007). 237 of the lakes in the watershed have measurements of mean DOC concentrations available (from the ILDB – Inland Lake Database - the mean values are based on different amounts of data). The 859 lakes cover 15 % of the surface area of the catchment. The average surface area of the lakes is ca 80 ha with a range of 5 to 12 000 ha and 60 % of the lakes are headwater lakes. About 10 % of the land area consists of wetlands.

The area lies in the southern Boreal Eco Climate Zone (hydrological data for the region see Table 2) of the Canadian Shield and the entire region is underlain by bedrock consisting of Precambrian metamorphic plutonic and volcanic silicate. The topography is varying with highlands, rocky knolls and ridges in the middle and lower parts of the watershed and these areas contain tiny sandy till. In the central parts there are some valleys, where there is deeper sand, silt and clay and these areas support farms with fields for pasture. The forests in the area are often dense and consist of mixed hardwood of maple, birch, and oak as well as coniferous species like spruce, white and red pine, balsam, fir, tamarack and hemlock. (O'Connor, 2007; www, MWCI, 2007)

Of the about 150 000 inhabitants in the catchment about two thirds are seasonal (O'Connor, 2007; www, MWCI, 2007). Many of the animals in the area have a life cycle that is related to the river and/or lake and the wetlands (www, MWCI, 2007).

Table 2. Hydrological data for the Muskoka River Watershed in Ontario, Canada.

	Value	Units
Average annual precipitation	1000	mm/y
of which is snowfall	300	mm/y
Long-term average catchment runoff	506	mm/y
Mean January temperature	-10	ºC
Mean July temperature	17.7	ºC

Source: (O'Connor, 2007; www, MWCI, 2007)

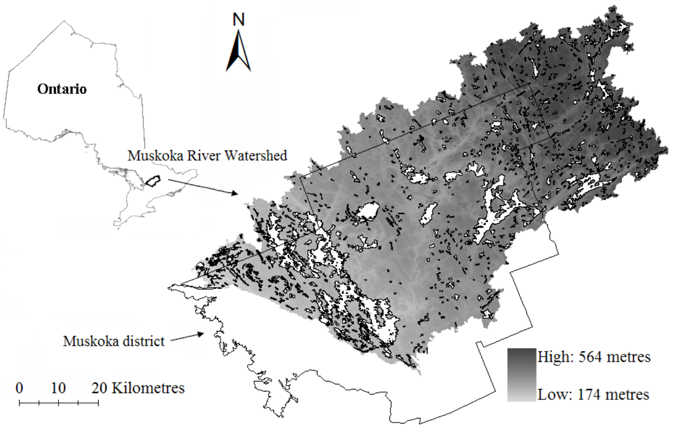


Figure 4. The location of the Muskoka River Watershed (MRW) in Ontario, Canada, and the watershed showing all the lakes of the watershed as well as the elevation. The outlined area overlaying the watershed is the Muskoka District (MD). (Picture made in ArcGIS 9.1 and Microsoft ® Paint 1.5.)

4 THEORY

Regression is to look back, in this case to use parameters for the catchments to explain the level of DOC in the river. Multiple regression, regression with more than one parameter, can be seen similarly as single regression, but with matrixes instead of a single parameter dataset. Multiple parameters also bring new problems and the need to look at the significance, not only of the model, but of each single model parameter as well. A multi-model approach is also a road more often taken by modelers in recent time, as more than one model might be possible to choose and the result between models compared, leading to more knowledge being obtained about the system.

Statistical analysis is what you use to build the model, see how good it predicts your data and how sensitive it is to changes in the in- and output used to calibrate and validate it. If many parameters are used the relationships between them become important as this points to the possibility that several parameters might explain the same parts of the output and may not all be necessary, or they may affect each other indirectly.

4.1 MODELS

Why one should model environmental substances (Schnoor, 1996):

- To gain better understanding of their fate, transportation and so on
- To determine concentrations in organism in the past, present and future.
- To predict the conditions in the future under different scenarios, for example the climate scenarios used in research at present.

To model aquatic chemicals four ingredients are needed (Schnoor, 1996):

1. Field data on chemical concentration and discharge: [DOC], Q
2. A mathematic model formulation: [DOC] =
3. Rate constants and equilibrium coefficients for the model: b_i
4. Some performance criteria with which to judge the model: r^2, p, F, …

Models can be built on known knowledge to gain new knowledge. One can analyze and explore scientific problems trying to identify what lies in the gaps of what is already known and try to identify the key processes, rates and parameters (O'Connor, 2007) in a system or for a certain substance. Models are simplifications of the reality, but can still explain and give knowledge of the problem/system at hand.

In this project new mass balance models for stream DOC are being built and then used in connection with another mass balance model, the export Lake DOC Model (LDM). Other models are also available to model the DOC, with different approaches and built in differing areas.

4.1.1 Mass balance models

In mass balance models a systems input and output of some substance is studied, which is a part of why it also is called input-output budgets (Evans *et al.,* 1997). If one flow exceeds the other the system is not at equilibrium or steady state (a common approximation is to assume that the substance or system is at steady state), but the system accumulates or loses mass of the substance. With mass balance models one can gain knowledge of the sources, sinks and other fates (for example; consumption, sedimentation) of the substance in question.

The advantages of doing mass balance modelling (Evans *et al.,* 1997):

1. They are holistic, information is integrated over a large region and long time
2. The computations are relatively simple
3. The base is the fundamental principle of conservation of mass
4. Routine monitoring programs can provide the necessary input data in some whole system studies

A box model (the box is the control volume of the model, the boundary of the area which the out- and inputs must cross, see for example Figure 5 below showing the models which will be used in this project) can be one approach to solving mass balance transport equations and with this approach simplifications can be made and ordinary differential equations be used instead of partial (Schnoor, 1996).

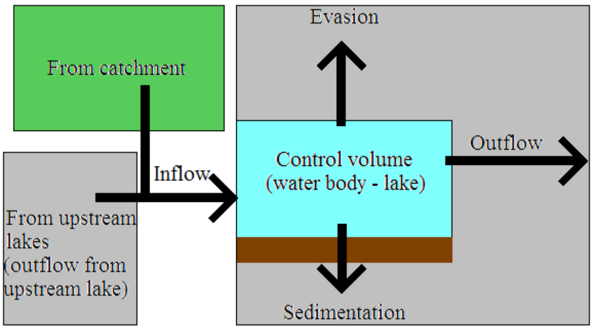

From catchment
From upstream lakes (outflow from upstream lake)
Inflow
Evasion
Control volume (water body - lake)
Outflow
Sedimentation

Figure 5. Example of a control volume and the different in- and outflows. The top left part is estimated with the mass balance model to be developed and the middle light grey parts are computed with the Lake DOC Model.

4.1.2 Lake DOC Model

The Lake DOC Model (LDM), used in this project to model the fate of DOC once it has entered one of the lakes in the catchment, was put forward by Molot and Dillon (1996) and Dillon and Molot (1997a). The Lake DOC Model is a rather simple model developed to examine the fate of DOC in lakes. Several assumptions were made, for example that the lakes on an annual basis were at steady state and that the outflow and losses could be seen as first-order kinetics. The LDM sees the input of DOC as mainly coming from the catchment via the stream and the flow of groundwater DOC directly to the lake was neglected as was production of DOC in the lake. The DOC was seen as leaving the lake mainly via the outflowing water, degassing (CO_2 entering the atmosphere after decomposition) and sedimentation (see section 2.3). See Figure 6 for a view of the in- and outputs from the model. Data from the seven lakes, described in section 3.1 were used in the development of the model. During the 20 year period used in making the model there were both dry years and wet years, but no overall trend towards a change in climate.

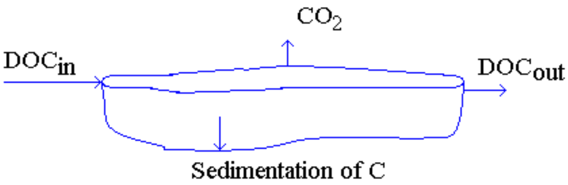


Figure 6. The in- and outputs from a lake regarding the Lake DOC Model.

A box model was used, but with only one box, which means that an assumption had to be made that in the whole lake the concentrations were the same, which is called a complete-mixed criteria. This leads to equations like:

$$z \cdot \Delta[DOC] = DOC_{in} - DOC_{out} - v \cdot [DOC] \qquad \textbf{(1)}$$

where z is mean depth, Δ[DOC] is the difference in mean concentration (successive years), flux of DOC in (DOC_{in}) and out (DOC_{out}) of the lake, v is the net generalized loss coefficient and [DOC] mean annual concentration.

The retention (R) of DOC in the lake (Molot and Dillon, 1997b) can be computed by different means according to the equations (values for the lakes are shown in Table 3 below):

$$R = \frac{DOC_{in} - DOC_{out}}{DOC_{in}} \quad \textbf{(2)} \qquad \frac{1}{R} = \frac{q_s}{v} + 1 \quad \textbf{(3)} \qquad R = \frac{L - L_o}{L} \quad \textbf{(4)}$$

where q_s = areal runoff, v = net loss coefficient, L_o = loss via outflow and L = load.

The net retention of DOC (R_{doc}) i.e. the total amount that did not discharge with the outflowing water was measured to be between 40 and 70 % in the seven lakes. The partitioning of this amount of DOC between degassing and sedimentation was seen as depending on the lake's alkalinity in such a way that there were more degassing relative to sedimentation with a decrease in alkalinity.

The Lake DOC model is used to gain, by determining the loss of DOC, the outflow of DOC from the lake. This is then entered as inflow of DOC to the next level lake, thereby connecting lakes in a watershed. For this study losses in streams for the flow coming from another lake will be neglected. The last term in the equation 1 was divided into two terms: the loss of DOC from upstream lakes and from the catchment. This gave the need for two v´s; one for DOC coming from upstream lakes (v_u) and one for DOC coming from the catchment of that lake (v_l). Both are affecting how DOC that enters the lake is exported from the lake and sedimented/evased. As a default value (in the actual model Excel spreadsheet used in this work, values also used in the work of O´Connor (2007)) these were the same, both 3 m/yr. According to O´Connor (2007) their ranges were: v_u: [0; 3] - loss coefficient for DOC from upstream lakes

v_l: [0; 6] - loss coefficient for DOC from the catchment of the lake

These ranges were not fixed because they where based on only the seven lakes in the Dorset study (personal communication Peter Dillon). For more information of the Excel sheet in which this model was used together with the models derived in this paper, see O'Connor (2007, chapter 3 and Appendix G).

Table 3. Data for the equation parameters for the seven Dorset lakes. The mean and standard deviations (sd) are given for DOC input and outputs during 1981 – 1989. v is computed from (3) and R from (4).

	q_s [m yr^{-1}]		R [-]		v [m yr^{-1}]	
Lake	mean	sd*	mean	sd	mean	sd
Blue Chalk	1.50	0.29	0.59	0.04	2.2	0.3
Chub	4.23	0.78	0.42	0.05	3.0	0.3
Crosson	5.60	1.09	0.37	0.05	3.3	0.6
Dickie	2.66	0.59	0.55	0.05	3.2	0.4
Harp	4.16	0.75	0.42	0.04	2.9	0.4
Plastic	2.00	0.38	0.69	0.03	4.6	1.0
Red Chalk	5.44	0.99	0.37	0.06	3.2	0.7

Sources: Dillon and Molot, 1997b

*sd = standard deviation

4.2 MULTIPLE REGRESSION

Multiple regressions involve more than one parameter and the first-order model (meaning linear in parameters and response) becomes:

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + ... + \beta_j \cdot x_{ij} + ... + \beta_p \cdot x_{ip} + \varepsilon_i \qquad \textbf{(5)}$$

where **p** is the number of parameters in the model, $\boldsymbol{\beta_i}$ are constants which are called both population parameters and regression coefficients and are attained through, for example, the method of least square (see section4.2.1). β_0 is the intercept of the model, while the rest of the β parameters are slopes for the different predictor variables (Quinn and Keough, 2006). **ε** is the random or unexplained error term and $\mathbf{x_{ip}}$ are called predictor variables or simply parameters (Neter *et al.,* 1989; Quinn and Keough, 2006)

The parameters can interact with each other, adding more terms (like $x_1 \cdot x_2$). This work does not involve the interaction terms, but the multiple regression equation above can be used for this case as well by calling $x_1 \cdot x_2 = x_{p+1}$. Even variants of parameters, like log (x_i), or x_i^2, can be renamed and entered into the model. Once the parameters have been chosen they can be fitted to a regression model:

$$\hat{y}_i = b_0 + b_1 \cdot x_{i1} + b_2 \cdot x_{i2} \ldots + b_j \cdot x_{ij} + \ldots + b_p \cdot x_{ip} \qquad \textbf{(6)}$$

where $\hat{y}_i$ is the simulated value of the modelled parameter and $\mathbf{b_i}$ is the estimate of $\boldsymbol{\beta}_\mathbf{i}$. The $\boldsymbol{\varepsilon}$ is not a term here, the error between estimated and actual values is called residuals and is the values to be minimized during the regression process.

4.2.1 Least Square and residuals

Ordinary Least Square (OLS) is the same for multiple linear regressions as for simple linear regression, but with multiple regressions there is more than one set of normal equations to solve, as there is one set for each of the parameters to be estimated (Quinn and Keough, 2006). The normal equations can be set in matrix form; more information on this is available in statistical literature, for example (Tabachnick and Fidell, 2007).

Least square analysis estimates the β_i´s to b_i´s by minimizing the error ($\boldsymbol{\varepsilon}$) between the observed and the predicted values squared (Eq 7), the Residual Mean Square:

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \qquad \textbf{(7)}$$

The residuals are also used to test the fit of the model to the data, as residual analysis is a way to see if the model is valid. The residuals need to be normally distributed for a linear regression to be valid. The residuals reveal if a major part of the y_i´s are not explained by the regression model, by not being normally distributed and exhibiting a trend. (Neter *et al.,* 1989; Tabachnick and Fidell, 2007).

To attain the regression equation (Eq 6), different methods of multiple least square can be used, and different statistical programs use different types. One factor that separates the types of regressions is in which way parameters not yet entered into the regression equation are weighed. The weight is often based on how much (more) of the variance they explain in the y. This is a way of letting only important parameters enter. With more than one parameter the variability that one of them explains in y can be divided between:

- Unrelated/unique – seen to the variability explained by the others
- Related - meaning that they also explain the same variability in y.

Some programs give the choice between different types, or let the user interact in the process, by choosing which parameter to enter next. There are three basic types of regressions: standard, sequential and statistical, which looks at different parts of the variability of parameters as a reason to enter a certain new parameter into an equation. The latter two are used in this project.

In sequential regression the researcher/user chooses the order in which the parameters are added to the regression and each one parameter is evaluated at the point of entry on what it adds in explained variability, meaning that all variability shared by parameters already in the regression equations is not counted and the order of entry is of major importance (Tabachnick and Fidell, 2007).

Statistical regression is usually called stepwise regression, even though there really is three different parts – forward selection (starts empty and has an entry criteria), backward deletion (starts out full and then has a removal criteria) and stepwise regression, that is a compromise between the other two, having both entry and removal criteria (stepwise forward and stepwise backward) (Tabachnick and Fidell, 2007). The statistical criteria that determine the point of entry or removal means that as in sequential regression the parameters are as important as the part of the variability not shared by another parameter that has already been entered. This might in affect turn out parameters almost as highly correlated as the first to be entered if a big part of its variability is eaten by other parameters that is if it is highly correlated also to the other parameters (Tabachnick and Fidell, 2007). For statistical regression cross validation is highly recommended, meaning that some data is used in calibration of the model and some data for validation of the model equation found.

4.2.2 Calibration and validation

Building a good model requires two different sets of data for the calibration and validation, taken at different locations or times (Schnoor, 1996).

Model calibration means finding parameters to the model that gives the least error between observed field data and simulated data from the model; this is where the regression equation is obtained.

The validation of the model involves a dataset not used for calibration and a comparison is made between this dataset and a simulated one from the model, with the x_{ij}´s associated with time and place. In this step the model is not altered in anyway to adjust the result (Schnoor, 1996).

The validation is also where the model can gain some scientific acceptance as to whether it contains all major, in this case, sources of DOC and if they are expressed in the right way (as linear relationship, negative or positive to [DOC]) and if it can describe the concentration of DOC as was its intention. Validation can also come from using the model in a wide range of areas where there is some sort of data to compare the result. This multiple testing is also a test of the models robustness and power.

4.3 RELATIONSHIPS FIT AND POWER

A relationship is based on a correlation between two parameters. The presence of other parameters can affect this relationship, by parameters being correlated among them selves (multicollinearity) and by one parameter suppressing the effect of another. The fit of a model depends on how well it can explain a dataset and its power lies in not being too sensitivity to changes in input or output.

When choosing which parameters to actually use in the final equation/model this must be taken into account, but also the choice will depend on knowledge of the parameters themselves. What correlation as well as regression does not take into account is the errors in retrieving the parameters or other facts as, in case of the model being used for prediction, the availability of the parameters. Sometimes a set of parameters might be chosen that leads to an equation with a lower estimating power than another due to anyone of these or other factors. One such factor can be the fact that a model with one less parameter can be a better choice if the gain of the last parameters is deemed too low in comparison to the worked need to attain the parameter.

4.3.1 Correlation and multicollinearity

For a regression equation to be of any importance a relationship between the x_i´s and y are of importance. If many x_i´s are to be used the correlation between these will also need to be taken under consideration as a high correlation between two x_i´s imply that using both will not add much to estimating y as they will estimate a similar part of y. The correlation between both y and x_i´s and the different x_i´s are best calculated with a correlation matrix. From this several variables with high correlation to y can be taken out and used in regression. Sometimes it might be of use to put the variables in groups if many variables are also correlated with each other. (Tabachnick and Fidell, 2007)

Correlation between parameters used in the regression can also lead to problems in the regressions process, something called multicollinearity. Multicollinearity means that there is a linear dependence (i.e. a relationship) between the parameter datasets of x (Vinod and Ullah, 1981). There are ways to located if this is a problem:

- ❑ Any eigenvalues of the matrix is close to zero.
- ❑ A VIF (variance inflation factors) > 5 means that the multicollinearity would be harmful (Vidon and Ullah, 1981)
- ❑ Do a sensitivity analysis (see section 4.3.5) on the model results and see how that affects the regression coefficients (variance of these should be low).

The problem arises as the inverse of the correlation matrix is being calculated during the regression process and with high correlation between parameters this can become singular or close to singular (singular meaning that the inverse of the matrix does not exist) (Vinod and Ullah, 1981; Tabachnick and Fidell, 2007). The variance of the regression coefficients is also increased with multicollinearity.

4.3.2 Suppressor variables

A suppressor variable is a variable that, when entered into a regression, affects the variance of another variable (x_i) already in it. The variance is suppressed, diminished, and this increases the r^2 value of the regression. The variance that is suppressed is the variance in x_i that is not relevant in predicting y. (Tabachnick and Fidell, 2007)

One sort of suppression is negative or net suppression in which the presence of a suppressor changes the sign of the regression weight, or regression coefficient, for one of the variables. The changes give the regression weight an opposite sign from what might be expected from its correlation with y and the sign that variable would get in a single parameter regression between that x and y. The variable that is suppressed might also have a stronger than expected effect on the y, i.e. a larger regression coefficient. (Tabachnick and Fidell, 2007)

Up till know there is no available test for finding suppression. It is also hard to find which variable is doing the suppression when the regression contains many parameters as the suppressor is not affected, but has a regression coefficient that has expected weight and sign. It might be found by leaving other variables out, one by one. If it can not be found it might be better to just look at the implications of the simple fact that the suppression exists. (Tabachnick and Fidell, 2007)

4.3.3 Importance of a model – R, F, t and confidence intervals

These values are normal output from most statistical programs when a multiple regression has been made and are a way of gaining knowledge of the power and fit of the regression.

For multiple regression R^2 (maybe to point out that matrix algebra is used) is often used above r^2 (and $R = \sqrt{R^2}$ above r) and an adjusted R^2 (R^2_a) is often computed, which adjust the R^2 by dividing each sum of squares making up the variable by its degrees of freedom (df). This makes the effect of adding more and more parameters into the equation, something that is not always positive, less straight forward as the adjusted R^2 does not always increase as R^2. These parameters all explain how much of the variability the regression explains; 1 means 100 % and 0 nothing. (Neter *et al.,* 1989)

Depending on the sample size the R-value holds differing power, which is especially important to think of with small sample sizes. Table 4 show the significance of the r values for the number of values that are used in this project (large values not added).

Confidence intervals for b_i´s can also be obtained from many programs. The coefficients are found to be significant at a certain level (95-% being the most common) if zero is outside of the interval.

Many programs also show the results of t-tests assessing the significances levels for the regression coefficients. This is related to the F (a parameter also usually obtained) with the relationship $F = t^2$ (Neter *et al.,* 1989).

Table 4. The significance of an r at certain level of different number of data points, n.

n	r (0.05)	r (0.02)	r (0.01)
5*	0.878	0.882	0.917
12	0.576	0.658	0.708
15*	0.514	0.592	0.641
16	0.479	0.574	0.623
17	0.482	0.558	0.606
19	0.456	0.529	0.575
20	0.444	0.516	0.561

Source: Freund, 1967.

* fifteen datasets are used for calibration and five for validation, during model development. The remaining n is the number of years of data available for any of the subcatchments in the Dorset study area.

4.3.4 Normality tests and nonparametric test

For a regression to actually be accurate and have a deterministic power outside of the dataset being used in calibration, the data, y, x_i´s and residuals need to be normally distributed (coming from a Gaussian distribution) or at least approximately normal (Neter *et al.,* 1989). A normality test usually does not have enough power when the sample is small to actually tell if it comes from a Gaussian distribution. Small samples usually pass the tests. When the sample sizes are larger even small deviations from normality may be flagged as significant. (www, GraphPad, 2007)

There are different normality tests available, all having negative and positive aspects:

- The Kolmogorov-Smirnov normality test checks the cumulative distribution of the data to the expected cumulative Gaussian distribution and bases the p-value, on which a dataset passes or fails the test, on the largest discrepancy. It is well known, but some believe that it might be too simple (www, GraphPad, 2007).

- The D´Agostino-Pearson omnibus test determines the skewness/asymmetry and the kurtosis (to quantify the shape of the distribution). The result is a single p-value from the results of the effect of these values. (www, GraphPad, 2007)

- Shapiro-Wilk normality test is also used, but it will have problems if values are not completely unique.

Some nonparametric tests can also be used to gain some insight into the normality of the data (www, GraphPad, 2007). For example:

- Wilcoxon signed Rank test that compares the median to a hypothetical median (here compared to a Gaussian distributions median).
- Unpaired t-test, compares the mean to the mean of a hypothetical mean (here compared to a Gaussian distributions mean).

4.3.5 Sensitivity and uncertainty analysis

Sensitivity analysis (SA) involves determining the effect that small changes in the model parameters have on the results (Schnoor, 1996). In short, to see how sensitive the output is to input (O'Connor, 2007). The result shows which input parameters need the most accurate measurements as their eventual errors will have the largest effect on the result. It can be made by multiplying one parameter set at the time in a model, with a random number between ± n %.

Uncertainty analysis involves determining uncertainty, or standard deviation, of the output from the expected mean, due to the model inputs with stochastic techniques. This is more the error due to output, not input. This is done during the estimation of the parameters in the regression equation.

4.4 MULTI-MODEL APPROACH

This is the way things are done nowadays. More and more modelers look at multi-models instead of just one model (personal communication Julien Aherne). In the study of Demtener *et al.* (2006) the result from 23 models and the average of these models were evaluated for nitrogen and sulfur deposition. One important result was that the average (or mean) model was among the best.

With multiple models estimating the same thing, it is possible to choose the best one for the intended purpose or the one for which there is input data available. It can also be useful to have several models and average the result from all of them or the n best models.

5 METHOD AND PERFORMANCE

At least one mass balance model to estimate the stream concentration of DOC is the goal of this project. Together with the Lake DOC Model (see section 4.1.2.) this will estimate the DOC of the whole Muskoka River Watershed, the same area as already used in the work of another Master Thesis at Trent University in Ontario (O'Connor, 2007). There the 1997 catchment model for stream DOC concentration, which uses a relationship between the DOC and wetland percentage, obtained from work performed by Dillon and Molot (1997b) was used. This project will continue their work, as more DOC data is available, as well as better GIS data.

The project was first focused on a literary review which looked into DOC and the possible catchment parameters to explain the flux of DOC. It also involved looking into trends and groupings (Brien's Test) for available DOC concentration and flux data in the Dorset study area (the 20 subcatchments used in the model building). Positive trends have been seen in other parts of the world, mainly Europe (Dillon and Molot, 2005) and the North America, while parts of Canada as well as North-eastern US and Germany have noticed negative trends (Futter, 2007).

The next part was looking for GIS data from which the possible parameters, found through the literary review, could be obtained. The data was applied for mainly through the Bata Library at Trent University and GIS layers for both areas were asked for at the same time, though they might not be needed for the larger area if the parameters did not end up in the final models. The reason for this was that the areas were close and somewhat overlapping and two of the three areas needed were needed for the smaller area.

The parameters were then obtained with ArcGIS for each of the 20 subcatchments, exported to Excel and then used to build new mass balance models to explain DOC flux/concentration. The new models found to explain the most DOC were then applied to the Muskoka River Watershed to see how it performed there. Again data for the input to the model was obtained from GIS. Then the loss coefficients in the Lake DOC Model were optimized for each model and different amounts of the lakes with measurements (for example only headwaters).

5.1 DOC DATA FOR THE DORSET AREA

Data for the Dorset study area over the concentration of DOC and also DIC and TOC (total carbon, seen as DIC + DOC) was available between the years of 1978-1998. Data between the years 1998-2006 were not available yet as the study began, but had been measured for most of the 20 streams. The 20 subcatchments had 12-20 years of continuous data under the given time period and runoff (Q) data was also available for all the 20 subcatchments during the same years as DOC data.

5.1.1 Checking for linear trends for Dorset data

The mean DOC would be used in the models, but possible trend are still of importance as they might suggest that climate change or changes in other parts of the environment, like acidification, could have affected the data. This might mean that the model mean will not be representative for the entire period. Possible linear trends were investigated for each subcatchment measured data-series of DOC, TOC (DOC + DIC), DOC/Q and TOC/Q data against time, by simply adding a linear trend line to a graph in Excel. The four r^2 values obtained for each subcatchment were investigated for significance based on the number of measurements each subcatchment had (see Table 4 in section 4.3.3).

5.1.2 Brien's test – grouping of data

Brien's test is a test that finds the groups within a number of dataset. It tells which datasets are correlated enough to be considered to belong to the same group. This test was performed separately on each of the five datasets:

1. Measured data-series a). DOC and b). TOC
2. Calculated a). DOC/Q and b). TOC/Q
3. Values of the parameters obtained from GIS.

The analysis was done with Excel with the correlation matrix between the datasets and imported Spreadsheets (see Appendix B for an example of the spreadsheet), in which the actual tests were performed. The plan was then to:

1. Find the two parameters that have the highest correlation in the correlation matrix (using only the lower or upper half and not the diagonal of the correlation matrix). These are the first parameters entered into a group and from now on called A and B.

2. Using the entire correlation matrix the correlations between all members of the group are average for each other parameter not already entered into any group. The one with the highest value is called C, D, … .
3. Now the actual Brien's Test is performed: first a BT3 (as three parameters are to be analysed, the spreadsheet for a BT3 analysis is shown in Appendix B, Figure App-8), BT4, BT5 and so on. Spreadsheets were available for BT3-BT11 and BT18 (could be used to gain BT12 to BT17 which were never needed). The main result used for the test was the p-value for Equal correlations:
   1. If $p > 0.05$ the new parameter is entered into the group. Back to 2, to find the next parameter to test.
   2. If the $p < 0.05$ the new parameter is not entered into the group and the group is full, no more parameters will be entered instead a new group will start if there is enough parameters left to start one. First the correlations between the members of this group are removed. If more correlations remain the highest correlation is looked for again under 1, finding new A and B's.

If two datasets/parameters remain outside of any other group in the end they are grouped together and if one dataset remain it will be alone in one "group".

At the top of each BT´n spreadsheet the number of parameters (n) and the number of years that the datasets have data needs to be entered. 20 years was used as more than half of the subcatchments have 20 years of data (personal communication, Joe Findeis). During one grouping procedure (all Brien's test performed on the same correlation matrix) the numbers of years must remain the same to not disturb the analysis.

5.1.3 Normality of measured data

Normality tests were performed in GraphPad, with data imported from Excel, where the analysis gave a new sheet with all the results from a range of tests that had been chosen.

5.2 DOC DATA FOR THE MUSKOKA RIVER WATERSHED

DOC data was available for the Muskoka River Watershed in the Excel spreadsheet model where the mass balance model of Dillon and Molot from 1997 was linked to the Lake DOC model. This sheet was used in this project too, but the model equation was altered to the ones found in the first part of the project. The DOC data was accessible for 237 lakes (117 of those were headwaters) and the data originated from the ILDB (Inland Lake Database) where different amounts of yearly data were averaged and entered into the model. These measurements were also used in the work by O'Connor (2007) and as the data were measured in different ways (mainly all year around or only seasonally, i.e. no measurement during ice coverage) an analysis was performed during the thesis work of O`Connor in which no significant difference in the mean DOC concentrations were found. Normality tests were also performed on this data within the program GraphPad.

5.3 POSSIBLE PARAMETERS

Possible GIS parameter, found from literature and personal communication with P Dillon:

- ❑ Drainage density = stream length/catchment area
- ❑ Catchment and stream slope (average)

- ❑ Wetland: percentage, classes for wetlands (based on length of the stream to the lake: average, max, min stream length for all wetlands as one, wetland with forest, open land, open water [ponds])
- ❑ Agriculture/pasture
- ❑ Forest, forest clear cut, forest types
- ❑ Open land/bedrock/soil/road
- ❑ Open water (more than the lake – ponds, …)

Parameters for which it might be hard finding data that is not too coarse: Cryptic wetlands (need LiDAR as it has 2.5 m accuracy), forest on the riparian zone within different distances (the usual accuracy of 30 m is not good enough), water residence time, flow paths, and elevation. Lake data will not be used since stream DOC, in inflowing water, is the parameter to be estimated.

This leads to the need of the following GIS layers (besides catchments and lakes that were already available): DEM, wetlands, streams, forest, agricultural, open land, pasture, soil, bedrock and roads.

5.3.1 GIS data

Most of the GIS data were available from Trent Bata Library through their agreement with MNR. Some data were also available from Ducks unlimited and some data earlier used had been altered by GIS technicians at Trent, like the delineation leading to catchment and lake layers. The road layer came from the Geographic network online (www, OBM, 2007). Data were needed for three areas; Bancroft, Parry Sound and Algonquin Park, the latest only for the MRW. DEM´s did not follow these areas, but came in even smaller parts and data was sent for the whole of zone 17 in Ontario.

The wetland data was available from two sources:

- The Natural Resource and Values Information System (NRVIS; from Ontario Ministry of Natural Resources (MNR)), which was given through the Bata Library (called OBM or NRVIS wetlands). This data covered all catchments in both areas.
- Ducks unlimited, wetlands attained with the Rapid Assessment Technique (RAT, therefore called RAT wetlands). More about the Rapid Assessment Technique can be read in O´Connor (2007, Chapter 3). This was a newer layer than the one used in that study and covered a larger area, meaning that the entire 20 subcatchments were covered and 756 of the 859 catchments of the MRW. The remaining catchments of the MRW were only partly covered or not covered at all by the layer.

A relationship was developed between the NRVIS and RAT wetlands from the 756 catchments that had both types of wetlands and this was used to fill the gap between them.

FRI (forest) data was available from the region districts office of Parry Sound and Bancroft and from the Bata Library. These layers also contained bedrock outcrops data. Stream data was available from Ducks unlimited through earlier studies. DEM, bedrock, soil and agriculture was available through the Bata Library. The road layer came as already mentioned from the geographic network.

5.3.2 Attaining parameters from GIS layers for the Dorset study area

How the parameters were attained from the GIS layers can be seen in Appendix H.

5.4 STATISTICAL ANALYSIS/MODEL BUILDING

Linear trends, correlations, groupings of data and multiple regressions as well as sensitivity analysis of the resulting models were performed.

5.4.1 Linear trends, correlation and multicollinearity

With the help of Excel and S-PLUS, simple linear regression was performed on parameters obtained from GIS. This was done against DOC, Q and DOC/Q, as well as between different parameters. A correlation matrix and Brien´s tests were performed in a similar manner as mentioned in section 5.1.1.

With the grouping of parameter data and the correlation between DOC and DOC/Q with the parameters as well as the parameters themselves high correlations as well as low could be identified. With some high correlations between x_i´s there was a risk of multicollinearity during multiple linear regressions.

5.4.2 Multiple linear regression

Multiple regressions were first performed on all parameters as well as a few subgroups with forwards, backwards and stepwise (i.e. statistical) regression, but also sequential regression (JUMP for example). Different programs were used (SPSS, Kyplot, Minitab, S-PLUS and JUMP), some of which could not perform regression on all parameters due to multicollinearity and some which could. SPSS was mostly used as it had a user friendly interface and gave results that were consistent with others, it was especially similar to S-PLUS (this was mostly used for single regressions and also later for multiple regression with less parameters). The regressions were made with all yearly values of DOC and runoff from each subcatchments, which gave a total of 384 values. This since all parameters could not be entered otherwise as they surpassed the number of subcatchments (26 to 20). Later only mean values were used (especially to gain the b_i´s actually used).

With the statistical regression F and p values were used to determine the entry and removal of parameters. In SPSS 15.0 the default values (others were also used) were:

Entry:	F = 3.84	p = 0.05
Removal	F = 2.71	p = 0.01

This gave a number of parameters that were mostly entered/not removed and a number of parameters that were often not entered, but removed. The choice of parameters was based on those that were not removed and often entered, together with their correlations to DOC and DOC/Q. One type of wetlands was also removed from further regressions, the one with the lower correlation to both DOC and DOC/Q.

The multiple regressions were then made with only 2-5 parameters and only on mean values of DOC. For this second round of multiple regressions a type of sequential regression, in which the user chooses which parameter to enter into the regression, was used. The choice was not based on order of entry but all parameters were entered at once and then a new regression could be made with less or more parameters. This took more time, but gave more control to the user and also allowed for entry of any subset of parameters of interest. For this S-PLUS was used.

From statistics data like r^2, eight models were chosen, which had 1-3 parameters (three of the models had two parameters and four had three) and the highest r^2 for that number of parameters. The parameters perRAT2 (explanation see list of Abbreviations) and perRAT3 were not used as they where so close to perRAT, which was more correlated to both DOC and DOC/Q (see Appendix G). A total of six different parameters were used in at least one of the eight models, average slope was used in all of the models (and was therefore always used as x_1, for convenience).

5.4.3 Multi-model approach, randomized calibration and validation

The data for DOC (y_1), DOC/Q (y_2) and each of the six parameters were divided into two dataset multiple times (10 000 runs were made). As now the mean values for each subcatchment was used, the total number of values were 20. Fifteen subcatchments were used for calibration and five for validation. The actual regression and selection of subcatchments were done in an Excel sheet. In the Excel file the catchments were chosen randomly and uniquely, for each model at the same time. This was done with the Excel function rand() (which gives an equal likelihood of a number in the middle of the interval as at the edges), changed to give an integer numbers between 1-20. The calibration datasets were chosen first and for each new dataset chosen it was checked against all others already chosen. The actual data was stored in a separate spreadsheet and imported to the sheet as the number 1-20 was coupled to the catchments (in the order of the subcatchments in Table App-1, in Appendix A), with the Excel function VLOOKUP(). With the extension called MCSim (www, IE, 2007) added to Excel this could be done any number of times. The extension uses Monte Carlo simulations and saves the selected data into a separate spreadsheet.

The regression was made from the first fifteen rows and the parameters b_0 (intercept) and b_1-b_3 (slopes for the parameters for each model) were obtained by the function LINEST in Excel: {=LINEST(Y;X´s;TRUE;TRUE)}
where Y stand for the column with DOC containing the first fifteen subcatchments, X´s stands for the column(s) containing the parameter values for the same subcatchments and the first TRUE stands for that the b_i´s are calculated normally (FALSE would give 0) and the second TRUE means that additional regression information (besides the b_0 and b_1 for a single linear regression) will be returned by the function.

The same function can be given too many cells, the more regression coefficients and statistical data that is needed the larger the output becomes. According to the Excel help the output array from LINEST is as seen in Figure 7. Some of the data, like regression coefficients and r^2 were exported for each simulation, as well as the r^2 for the validation dataset (it was computed with the function RSQ()) and the numbers (1-20) of the fifteen subcatchments that were used in that calibration.

In the 10 000 runs some duplicates were produced and these were removed with a program written in VBA (Visual Basic for Applications). For the code to work the data needed to be sorted and rounded (the original data was saved). The code then printed the unique datasets in a new spreadsheet. Different amounts of remaining decimals (5, 7 and 10) were used to round the data. Some differences occurred but in the end the result when using seven decimals was used.

	A	B	…			
1	b_m	b_{m-1}	…	b_2	b_1	b_0
2	se_m	se_{m-1}	...	se_2	se_1	se_0
3	r^2	se_y				
4	F	d_f				
5	ss_{reg}	ss_{resid}				

Figure 7. The output array from Excels function LINEST. In this model up to b_3 were needed, meaning that four columns in the first and second rows were needed.

5.4.4 Sensitivity analysis

The sensitivity of each model to each of its parameters were tested by altering one parameter at the time by a random number (using rand()) between ± n % (n equal to 10 and 25 were used) from the old values. All 20 subcatchment were used in an Excel spreadsheet and the output of the estimated DOC from 10 000 runs from the models were put in a separate spreadsheet with MCSim. The b_i´s used were a mean of all runs that gave an r^2 for both calibration and validations above 0.75. The random number was attached to one parameter at the time, meaning that for the models with three parameters MCSim needed to be run three times. One other thing was done, and that was to prevent parameters that represented percentages to gain a value above 100%.

5.5 DOC ESTIMATION FOR THE MRW

Three models were chosen to be used to estimate DOC for the Muskoka River Watershed. The old (peat) mass balance model from 1997 was also used, for comparison, to estimate DOC as the earlier study on the area did not use the new wetland data. The Excel spreadsheet with the original model was copied for each of the other three models and altered, mainly entering new catchment and lake areas and the new parameters, as well as changing the equation in the model column (for the old model new wetland, catchment and lake data were also entered). The different columns in the Excel spreadsheet are explained in Appendix F.

Once the parameters had been obtained in GIS they were linked to a lake attribute called MasterID and in Excel the data were matched with MasterID in the Excel spreadsheet model with the function VLOOKUP. The 237 of the 859 lakes that had measured DOC values were used to examine the results of the models (also subsets of these lakes, like only 117 headwater lakes, or for the models with wetlands all or only headwater lakes with RAT wetlands were used).

Then the other model in the Excel spreadsheet, the Lake DOC Model, was optimized for the old peat model as well as the three new once. The parameters optimized were the loss coefficients (v_u and v_l) as these had been derived from only seven lakes. This was also done with different amounts of the lakes with measurements.

5.5.1 GIS for Muskoka River Watershed

The work to attain the three parameters needed for Muskoka River Watershed can be seen in Appendix H.

5.5.2 Optimization of Lake DOC Model coefficients for the MRW

As the parameters were entered into the Excel spreadsheet results were immediately produced and could be analysed. Measured vs estimated could be plotted and residuals (measured - estimated) as well. The result from the residuals and the fact that default values (3) already entered into the model spreadsheet were used for v_u and v_l lead to a trial to optimize the Lake DOC Model, via the loss coefficients. In this optimization the ranges set for the parameters were ignored as they did not have a firm foundation. As a first step of optimization, to see in what direction of change the average of absolute deviation would diminish, the coefficients were both altered at the same time as well as one at the time to two and four (i.e. ±1).

With the extension MCSim one or both parameters were then changed within a range of values 1 000 times. For each run, the value of v_l, v_u or both as well as the value of the absolute average of deviations (the absolute value of each deviation of the estimate from measured value for the 237 lakes were calculated and then the averages were computed from these), and printed into a separate spreadsheet. This was done for the 237 lakes, as well as only the 117 headwaters and for models M3 and M8 all 175 lakes with NRVIS wetlands and only those that were also headwaters (90). The range was increased a few times as in most runs the minimum of deviations was found for values of the loss coefficients in the higher part of the range.

One problem was that with high loss coefficients most of the DOC is sedimented or evased, so DOC_{est} was low. Most residuals on the other hand were high (see a plot of 30 randomly chosen lakes in Figure App-13 in Appendix L) and the max deviations as well. The problem could not be solved in Excel and instead the program Crystal Ball 7 was used for the rest of the optimization. The ranges used for the optimization here was though found from the result from Excel, where max and min were computed after the runs leading to low r^2 (< 0.45) were removed. This gave ranges of v_u = [2, 20] and v_l = [2, 10] (except for model M8 where both max and min were above ten for v_l with Ducks unlimited wetlands only). For each model the number of optimization step were set to 2000 times and the goal used was to minimizing the average of absolute deviations, as before. As most optimizations gave value not at any end of the ranges they were not increased further. The optimized values of the loss coefficients were then entered into each model spreadsheet and new results produced.

6 RESULTS

Looking at the DOC data for the Dorset area gave some significant linear trends and subcatchment groups. With ArcGIS values for 26 parameters were attained for which linear trends, correlations, and statistics from multiple regressions gave a subset of parameters that explained the most of the flux/concentration of DOC. Three mass balance models were attained from the multi-model, multiple regressions with calibration and validation and sensitivity analysis. These, as well as the older peat model were used, together with the Lake DOC Model, on the catchments in the Muskoka River Watershed. Optimized values for the Lake DOC Models loss coefficients were found for the minimum average absolute deviation for each separate model, and subsets of data (all 237 lakes with measured data, only headwaters lakes which were 117, with Ducks unlimited wetlands coverage, all 175 lakes and among those, 90 headwaters lakes). Around 50 % of the DOC was explained by each model, but each model´s residual dataset also showed a similar bias.

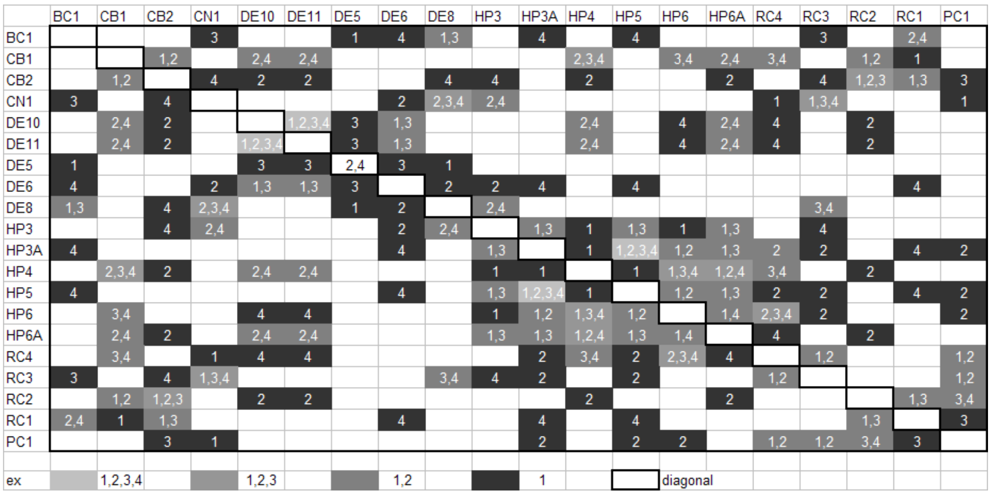
6.1 DOC DATA FOR THE DORSET AREA

The main results here were that some catchments showed a significant negative linear trend in the measured DOC data during the 12-20 years of measurements, but most trend were not significant. These were both negative and positive. From Brien´s test some groupings were found, in which some similarities were found, as the fact that all groupings produced five groups. For example the subcatchment HP5 was always grouped with, another subcatchment of Harp Lake, HP3A and DE11 were always in the same group as DE10, which both belonged to Dickie Lake. Other similarities were not always shared between all four sets of groups. The datasets being grouped together three times out of four did not always have the same connection to a lake. For the full grouping of datasets see Figure 8.

One other subcatchment to Dickie Lake, DE5, was found to be alone in two out of four cases. The linear trends that were made for yearly values of DOC, TOC, DOC/Q and

TOC/Q for each subcatchment showed that all the four highest r^2 values were also found in the same subcatchment (DE5). Its highest r^2 value, 0.6604, was for DOC divided by runoff. The lowest values of r^2 were more scattered between different subcatchments and the absolute lowest value was 3.00E-8 and it was for TOC in HP3, a subcatchment of Harp Lake.

The mean value of all r^2 values was 0.105. Even though the mean r^2 was so low 10 out of the 80 linear trends were significant, all of which were negative. The four linear trends for the DE5 catchments were significant at the highest level of probability, 0.01, as well as two of the other 10. One of the ten was only significant at the lowest level, 0.05. The other significant trends were TOC and DOC for DE6 (another subcatchment of Dickie Lake), TOC for RC3 (subcatchment of Red Chalk Lake) and TOC/Q for the subcatchments HP6, HP6A (subcatchments of Harp Lake) and RC2 (subcatchment of Red Chalk Lake). The negative trends are consistent with those found in the parts of Canada, the US and Germany, even though positive trends are more common (Futter, 2007). As mentioned positive trends has been found in parts of Canada as well.



	BC1	CB1	CB2	CN1	DE10	DE11	DE5	DE6	DE8	HP3	HP3A	HP4	HP5	HP6	HP6A	RC4	RC3	RC2	RC1	PC1
BC1				3			1	4	1,3		4		4				3		2,4	
CB1			1,2		2,4	2,4						2,3,4		3,4	2,4	3,4		1,2	1	
CB2		1,2		4	2	2			4	4		2			2		4	1,2,3	1,3	3
CN1	3		4					2	2,3,4	2,4						1	1,3,4			1
DE10		2,4	2			1,2,3,4	3	1,3				2,4		4	2,4	4		2		
DE11		2,4	2		1,2,3,4		3	1,3				2,4		4	2,4	4		2		
DE5	1				3	3	2,4	3	1											
DE6	4			2	1,3	1,3	3		2	2	4		4						4	
DE8	1,3		4	2,3,4			1	2		2,4							3,4			
HP3			4	2,4				2	2,4		1,3	1	1,3	1	1,3		4			
HP3A	4							4		1,3		1	1,2,3,4	1,2	1,3	2	2		4	2
HP4		2,3,4	2		2,4	2,4				1	1		1	1,3,4	1,2,4	3,4		2		
HP5	4							4		1,3	1,2,3,4	1		1,2	1,3	2	2		4	2
HP6		3,4			4	4				1	1,2	1,3,4	1,2		1,4	2,3,4	2			2
HP6A		2,4	2		2,4	2,4				1,3	1,3	1,2,4	1,3	1,4		4		2		
RC4		3,4		1	4	4					2	3,4	2	2,3,4	4		1,2			1,2
RC3	3		4	1,3,4					3,4	4	2		2			1,2				1,2
RC2		1,2	1,2,3		2	2						2			2				1,3	3,4
RC1	2,4	1	1,3					4			4		4					1,3		3
PC1			3	1							2		2	2		1,2	1,2	3,4	3	
ex	░	1,2,3,4		▒	1,2,3		▓	1,2		■	1		□	diagonal						

Figure 8. Groupings for the four different sets of datasets – DOC (1), TOC (3), DOC/runoff (2) and TOC/runoff (4), where the different colours (see bottom of picture) symbolize different number of groups the two datasets are grouped together in. The matrix is symmetric around its diagonal.

6.2 POSSIBLE PARAMETERS

From the data attained in GIS, 26 parameters of catchment properties were found. Linear trends, correlations, groups attained with Brien´s test and normality test were performed as well as a number of multiple regressions (with different programs). These showed that multicollinearity and suppression were present in different subsets of data. One statistical program could, for example, not perform multiple regressions on the whole set of parameters. As all parameters would not be used in one single model this was not seen as a too severe problem, but something to take under consideration. The end result was eight, one to three parameter models, used then in section 6.3.

6.2.1 Correlations, suppression and multicollinearity between parameters

The correlation matrix for all parameters and DOC as well as DOC/Q can be seen in Appendix G, Table App-2. The correlations were similar to the single linear trends attained for the parameters against both DOC and DOC/Q. In Table 5 a list of parameters with negative as well as positive correlation/relationship towards DOC and DOC/Q is shown. In Table App-4 in Appendix H the range, mean and other data is shown for each parameter in the Dorset study area (as well as for DOC and the parameters attained for Muskoka River Watershed).

During the different multiple regressions multicollinearity was sometimes a problem, and as the correlations between some parameters were quite high this was not a surprise. Suppression was also found, especially when yearly values and not mean values were used for DOC. As the number of parameters used were held below five, three suppressed variables were found and also a suspected culprit, average slope (to be observed is that at that time all parameters were no longer used). The suppression was found as some of the parameters changed signs (see section 4.3.2) in models with multiple parameters, compared to single linear regressions (see Table 5). The parameters that changed signs were:

- perFOR (percentage forest on catchments), changed signs in models together with: average slope alone and also as third parameter, percentage forest on RAT wetlands
- perFORRAT (percentage forest on RAT wetlands), changed signs in models together with: average slope and percentage forest
- strslope (Average stream slope), changed signs when together with: average slope, percentage RAT wetlands, and drainage density, alone and together

The most logical conclusion seems to be that at least average slope is a suppressing parameter and that average stream slope is the most sensitive to suppression. This was one reason that “strslope” was not chosen as a parameter in the last models, see section 6.3, as it otherwise also showed a high r^2 value together with for example average slope and percentage RAT (i.e. Ducks unlimited) wetlands.

Table 5. List of parameters based on their negative or positive correlation to DOC and DOC/Q, see also the correlation matrix in Appendix G.

Correlation to DOC and DOC/Q	
+	-
Catchment perimeter	Average slope
Catchment area/perimeter	Catchment area
Percentage OBM wetland	Percentage forest
Percentage RAT wetland	Small lakes/ponds (spond)
Percentage forest on OBM wetland	Small lakes/ponds (wpond)
Percentage forest on RAT wetland	Percentage OBM wetland 3
Straight distance to lake – OBM, average, max, min	Area of lake/area of catchment
Straight distance to lake – RAT, average, max, min	Drainage density
Percentage OBM wetland 2	Average stream slope
Percentage RAT wetland 2 and 3	Average stream slope/stream length
Percentage road length	

6.2.2 Brien’s test – grouping of data

There were nine groups found among the parameters (shown in Table App-3 in Appendix G). One of the groups contained the DOC, DOC/Q and the peat values used to obtain the 1997 mass balance model, indicating that this parameter is more related to the DOC than any of the new parameters. This was one reason why the 1997 model was also run with the new RAT wetland data in a later step to see if it could actually explain more of the DOC with new GIS data than the models developed in this study. Group eight contains two of the parameters found in the three parameter model developed and also used on Muskoka River Watershed.

6.2.3 Normality test on parameter datasets

Many of the parameter datasets failed some or all three of the normality tests, but many were ranked as Gaussian approximations by the Singed Rank test. The full table from the analysis of normality is shown in Table App-5 and 6, in Appendix H. This also contains the datasets for the Muskoka River Watershed (DOC measured and the parameters used in the models chosen in section 6.3).

6.2.4 Multiple regression

The multiple regressions gave a range of results; the main result was eight models having one to three parameters, using a total of six parameters. These had the highest r^2, and only models with r above 0.70 were chosen. The chosen models are shown in Table 6 below and the ranges of the six parameters are shown in Table 7.

Table 6. The models obtained from multiple regressions.

	x1	x2		x1	x2	x3
M1	Avslope		M5	Avslope	Wpond	PerFORRAT
M2	Avslope	Wpond	M6	Avslope	PerRAT	PerFORRAT
M3	Avslope	PerRAT	M7	Avslope	PerRAT	Wpond
M4	Avslope	PerFOR	M8	Avslope	PerRAT	Drainden

Table 7. The range, mean and median values of the datasets for the six parameter used in one or all of the eight models.

Parameters	Abbreviations	Min	Mean	Median	Max
Average slope	avslope	2.15	6.47	6.32	11.02
Percentage RAT wetland	perRAT	0	9.66	8.85	37.69
Percentage forest	perFOR	9.21	94.70	92.47	293.44
Percentage forest on RAT	perFORRAT	0	47.85	47.88	100
Ponds/small lakes	wpond	0	1.82	0	8.11
Drainage density	drainden	0	1.62E-3	1.66E-3	4.82E-3

6.3 MULTI-MODEL APPROACH

The eight models were used for a set of runs of calibration/validation of the models in Excel. From those mean coefficients were attained and then used in the sensitivity analysis for the models. From the results of calibration/validation and sensitivity analysis the resulting models were selected to be used in the Muskoka River Watershed.

6.3.1 Calibration and validation - uncertainty analysis

The set of 10 000 runs gave a high number of duplicates, how many for each model is shown in Table App-7, in Appendix I. Here is also shown the number of runs, all together and unique that gave an r^2 for both calibration and validation above 0.75. Table App-8 in Appendix I show which fifteen subcatchments were mostly used when high r^2 was attained.

The max and min value of regression coefficients for all runs and runs with $r^2 > 0.75$ and the mean and median for the ones above 0.75 are also shown in Appendix I. The resulting mean regression slopes for the runs $r^2 > 0.75$, were then used in sensitivity analysis, the results from which were used to draw conclusions on which (three) models would be used on the MRW.

6.3.2 Sensitivity Analysis

The sensitivity analysis was made with two different percentage changes (10 and 25 %) for the parameter values. The result from the 10 000 simulation runs of DOC values is shown in Appendix J, Table App-12-13 and Figure App-10). The largest ranges for the mean DOC of all the 20 subcatchments estimated, occurred for the parameter average slope. Some of the other parameters gave ranges for the estimated DOC that did not even contain the actual mean DOC.

6.3.3 Resulting models

The sensitivity analysis and the statistics from the calibration/validation resulted in three models; M1, M3 and M8 being chosen. One adding factor was that they had a natural progression as one more parameter is being added into each model, as M1 contains average slope, M3 is M1 + percentage RAT wetland and M8 is M3 + drainage density. This was one reason that model M8 was also chosen even though the range of the third parameter actually contains zero, even for runs with $r^2 > 0.75$. All other models also had zero in the range for at least one model parameter (see Table App-9, Appendix I). Models with the same x_i´s but used to estimate the variable DOC/Q are available in Appendix M, Table App-16-17. Equations were formed with the mean model parameters, from runs with an r^2 above 0.75, with four significant numbers (see also Table 8):

M1: DOC = 10080 - 669.9 · avslope (average slope)

M3: DOC = 9712 – 670.0 · avslope + 22.15 · perRAT (% wetland)

M8: DOC = 9711 – 611.5 · avslope + 23.32 perRAT – 3.110 10^5 · drainden (drainage density)

For statistical data for each model, see Table 8 that gives the r^2 and p. Otherwise, the results from the mean of the unique runs $r^2 > 0.75$ are available in Appendix I, Table App-10 and 11. Residual plots for the Dorset subcatchment for the three models are shown in Appendix K, Figure App-11. Due to the low number of data points no real conclusion can be drawn on the normality of residuals based on these plots.

Table 8. Values for the r^2, p and r when all 20 catchments are used with the mean parameters and when fifteen are used (five then separate)

		Calibration		Validation **	
Model	Equation	r^2	p	r^2	p
	All 20 catchments used in calculation of r^2 and p				
M1	10077.4 – 669.91 · x1	0.519	0.480		
M3	9711.99 - 669.98 · x1 + 22.1484 · x2	0.565	0.462		
M8	9711.27 - 611.51 · x1 + 23.3158 · x2 - 310975 · x3	0.570	0.461		
	15 catchments used for calibration and five for validation*				
M1	10077.4 – 669.91 · x1	0.786	0.391	0.826	0.430
M3	9711.99 - 669.98 · x1 + 22.1484 · x2	0.787	0.392	0.971	0.428
M8	9711.27 - 611.51 · x1 + 23.3158 · x2 - 310975 · x3	0.837	0.380	0.981	0.503

* The fifteen subcatchments used were the fifteen most often used in the 28 unique runs for M8 (used in at least seventeen runs) and the five runs unique for M3 (at least four runs) and the only run for M1.

** For validation n is only five compared to n = 15 for calibration. Smaller n can give a higher r^2, but it also needs to be higher to be significant (see Table 4).

6.4 MUSKOKA RIVER WATERSHED

DOC data for the Muskoka River Watershed, available in an Excel Model Spreadsheet was plotted for the 237 lakes in groups of low to very high concentrations. The DOC data was also used to gain residuals and to evaluate the result of each model. The GIS parameters were attained and evaluated and a linear relationship was computed between the different wetland types as RAT (i.e. Ducks unlimited) wetland data had a gap in the northern part of the MRW. The models were then used together with the Lake DOC Model to estimate the lake DOC. The result was evaluated based on the 237 lakes with data and the Lake DOC Models loss coefficients were optimized to attain a better result.

6.4.1 DOC data for the Muskoka River Watershed

237 lakes have measurements; the distribution of these data between low (< 3 mg/l), medium (3-6 mg/l), high (6-11 mg/l) and very high (> 11 mg/l) are shown in Figure 9 below (levels same as in O´Connor, 2007). The areas in the three circles are the areas that differs the most between the different models.

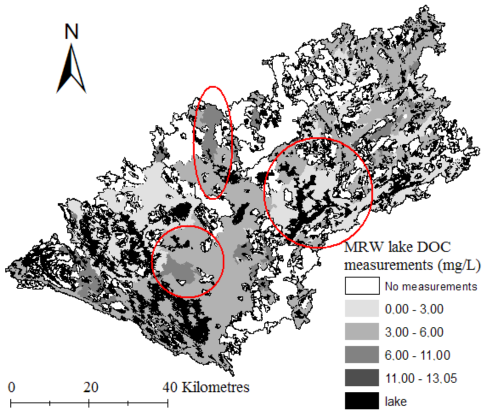


Figure 9. Dissolved organic carbon (DOC) measured for 237 of the 859 catchments of the Muskoka River watershed. The lake concentrations are here represented by the color of the catchment.

6.4.2 Parameters and wetland types relationship

The max, min, mean, median, standard deviations and other statistics as well as results of normality test on the parameter datasets are shown in Appendix H. Average slope was attained for all catchments. Streams were not present in 26 of the 859 catchments in the MRW, with the layer used. The RAT wetland layer had a gap in the northern part of the watershed, leaving 103 catchments totally or partially outside of the layer. The 756 catchments that had coverage were used to obtain a linear relationship between the wetland types, as NRVIS/OBM wetland data covered the whole area.

A pure linear trend was used, but linear trends with intercept zero, log and square of the percentage were computed as well. They all gave lower r^2 as results (0.1525, 0.1524 and 0.3435). Linear trends, with and without a zero intercept, for only headwater lakes gave r^2 of 0.2674 and 0.2859. Normality tests were performed with Prism on several of these wetland datasets computed for the linear trends and all datasets failed all normality tests but were seen as Gaussian approximations in the Signed Rank Test. Due to the high number of data (N = 756 or 460) even small deviations can result in a failed test and it does not have to mean that the data is not actually normally distributed.

The relationship used is: $\text{perRAT} = 6.184 + 0.9101 \cdot \text{perOBM}$
the relationship has an $r^2 = 0.3832$ (Excel and S-PLUS), RSME = 8.1040, F = 468.4, and correlation = - 0.5995.

This relationship was added into the Excel spreadsheets for models M3, M8 and the 1997 peat model, on the 103 catchments in the gap.

6.4.3 Residuals for MRW and optimizations of the Lake DOC Model

The results from the first set of runs in Excel, where v_u and v_l were set to two, three or four is shown in Appendix L, Table App-14 and 15 and Figure App-12. In Appendix L (Figure App-13), is also a residual plot for 30 random catchments with values from Excels 1000 runs sorted for the minimum absolute deviations. The plot shows how almost all of the residuals are positive. In Figure App-14 a residual plot for M1 and of all 237 lakes is shown using the loss coefficients from the highest r^2 obtained. Here most of the residuals are negative. The result for the three different models, sorted for max r^2, gave the ranges r^2 [0.505; 0.538], absolute deviations [0.501; 0.986], v_u [3.349; 4.426], and v_l [2.013; 2,082],

The optimization in Crystal Ball 7 gave different optimized loss coefficients for each model; the result is shown in Table 9 below, and the ranges found were v_l = [4.798; 6.446] and v_u = [6.145; 20]. Two models gave a v_u of 20 (both cases for lakes with RAT wetlands only), otherwise the values were at some distance from the upper part of the range. The residual plots for the different cases for the models (all, headwaters and/or with RAT wetland coverage), with the optimized values from Crystal Ball are also plotted in Appendix L (Figure App-15 to 17).

With the default values of the two loss coefficients the distribution between positive and negative values was not good and a bias, a negative trend, was apparent. After the optimization the distribution was better but still somewhat negative as the average deviations were negative, but the bias remained. Graphs for the residuals are shown in Appendix L, Figure App-12 to 17. The bias was similar for all four models.

Table 9. Optimized results from the different 2000 runs in Crystal Ball 7.

	All lakes					Headwater lakes only			
Model	N	v_u	v_l	r^2	Abs dev	N	v_l	r^2	Abs dev
M1	237	8.727	6.123	0.453	0.452	117	5.911	0.451	0.405
M3	237	7.795	5.926	0.468	0.447	117	5.798	0.475	0.408
M3_RAT	175	20	6.446	0.460	0.263	90	5.798	0.460	0.255
M8	237	6.145	5.487	0.492	0.414	117	4.798	0.492	0.367
M8_RAT	175	20	5.525	0.475	0.243	90	5.340	0.468	0.222
Old	237	4.761	6.461	0.443	0.557	117	5.315	0.523	0.534
Old RAT	175	11.33	5.562	0.458	0.367	90	5.315	0.531	0.694

* Range found for all lakes and headwater lakes jointly, not for the old peat model.

** Maximum r^2 found from 1000 runs in Excel, with range 2-18 for both loss coefficients

6.4.4 Modeled concentrations of DOC in Muskoka River Watershed

Concentrations of DOC were estimated for the optimized values of v_u and v_l obtained for each model. The residual plots show that there is a bias in the residuals and there is also somewhat of an underestimation, except for model M8. After optimization r^2 was lower (see Table 10 below) compared to the default values of the loss coefficients (see Table App-15 in Appendix L) values. The models explained M1: 45 %, M3: 46-47 %, M8 47-49 %, and the old peat model 44-53 %, of the DOC in the lakes with measured DOC levels. The mean model based on the average estimates of the three developed models (called AM3) explained 47.4 % for all the 237 lakes and for headwater lakes. With the 1997 peat model also in the average model (called AM4) 49.7 % of all 237 lakes and 53.8 % of the DOC in the 117 headwater lakes were explained.

Figures 10-13 show the estimates with the same groups as for the measured (Figure 9) in the 237 lakes with measurements; in Appendix L all 859 lakes are plotted in Figure App 18 and 19. Evaluation of the results can be seen in Figures 14-16 and in Appendix L, Figure App-20 to 23 (all cases only for M3 as an illustration of the fact that as the lakes being evaluated is diminishes in numbers it gets harder to use the graphs to evaluate the results). In the figures an estimate was considered as good if it was ± 25% of measurement, below 75% as too low and above 125% as too high. Histograms over the percentage of the estimated to measurement are shown in Appendix L, Figure App 24 to 27 and the percentage in each category above is presented in Table 11, below. In this M8 seems to be the best as it has the highest amount of good estimates and the most even distribution, especially for all the 237 lakes. Based on percentage explained AM4 is the best.

Table 10. Results, regression coefficients, r^2, r and other statistics from comparison to measured data and the estimated. Based on optimized v_u and v_l for each case.

Model	N	Equation	r^2	r
M1	237	$DOC_m = 1.791 + 0.643\ DOC_{est}$	0.4533	0.67
M1 hw*	117	$DOC_m = 1.803 + 0.672\ DOC_{est}$	0.451	0.67
M3	237	$DOC_m = 1.784 + 0.645\ DOC_{est}$	0.4735	0.69
M3 hw	117	$DOC_m = 1.804 + 0.685\ DOC_{est}$	0.4746	0.69
M3_RAT	175	$DOC_m = 2.033 + 0.675\ DOC_{est}$	0.4598	0.68
M3_RAT hw	90	$DOC_m = 1.920 + 0.682\ DOC_{est}$	0.4599	0.68
M8	237	$DOC_m = 1.548 + 0.675\ DOC_{est}$	0.4924	0.70
M8 hw	117	$DOC_m = 1.4574 + 0.701\ DOC_{est}$	0.4924	0.70
M8_RAT	175	$DOC_m = 1.812 + 0.693\ DOC_{est}$	0.4753	0.69
M8_RAT hw	90	$DOC_m = 1.685 + 0.713\ DOC_{est}$	0.4684	0.68
Old**	237	$DOC_m = 0.492 + 2.506\ DOC_{est}$	0.4429	0.67
Old hw	117	$DOC_m = 0.563 + 2.377\ DOC_{est}$	0.5234	0.72
Old_RAT	175	$DOC_m = 0.491 + 2.643\ DOC_{est}$	0.4574	0.68
Old_RAT hw	90	$DOC_m = 0.551 + 2.517\ DOC_{est}$	0.5306	0.73

* hw stands for headwaters only

** Old means the old mass balance model from Dillon and Molot (1997b).

Table 11. The percentage of estimates in the categories good (± 25% from measured), too high (> 125 %) and too low (< 75 %), for each model and case.

	N		Percent too low		Percent Good		Percent too high	
Model	all	hw	All	hw	all	hw	all	hw
M1	237	117	17.72	15.38	52.74	47.86	29.54	36.75
M3	237	117	16.88	11.97	53.16	49.57	29.96	38.46
M3_RAT	175	90	10.86	12.22	45.71	51.11	43.43	36.67
M8	237	117	21.52	18.80	57.38	56.41	21.10	24.79
M8_RAT	175	90	14.86	16.67	48.57	51.11	36.57	32.22
Old	237	117	18.57	17.09	42.19	33.33	39.24	49.57
Old_RAT	175	90	16.57	15.56	37.71	31.11	45.71	53.33
AM3*	237	117	18.99	15.38	52.74	53.85	28.27	30.77
AM4**	237	117	16.88	12.82	55.27	51.28	27.85	35.90

* AM3 – is an Average Model, based on the three models developed here

** AM4 – is an Average Model, based on the three models + the old peat model

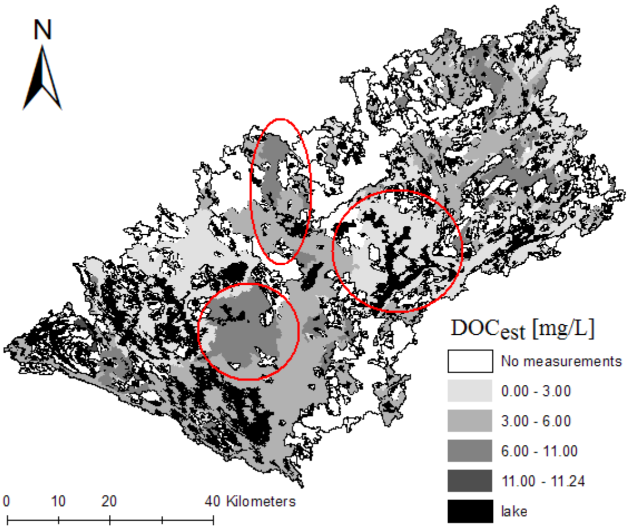


Figure 10. Estimated values for the 237 lakes with measured values for model M1. Look especially for differences between models and between measured values in the areas marked with a circle.

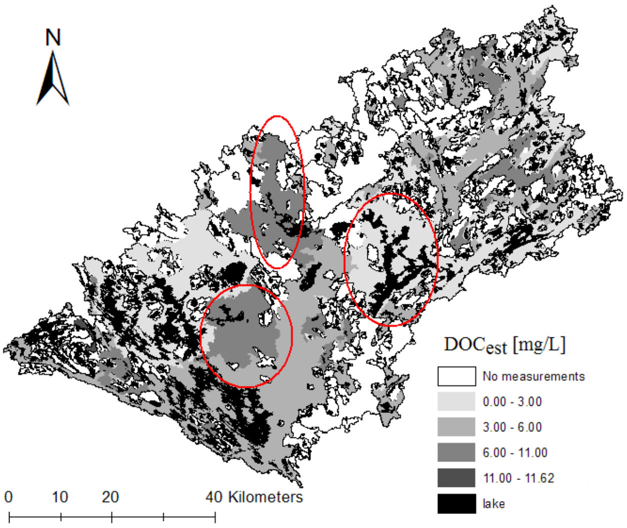


Figure 11. Estimated values for the 237 lakes with measured values for model M3. Look especially for differences between models and between measured values in the areas marked with a circle.

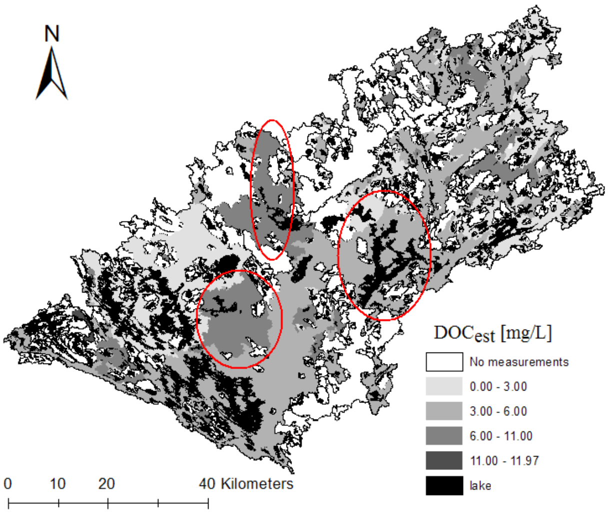


Figure 12. Estimated values for the 237 lakes with measured values for model M8. Look especially for differences between models and between measured values in the areas marked with a circle.

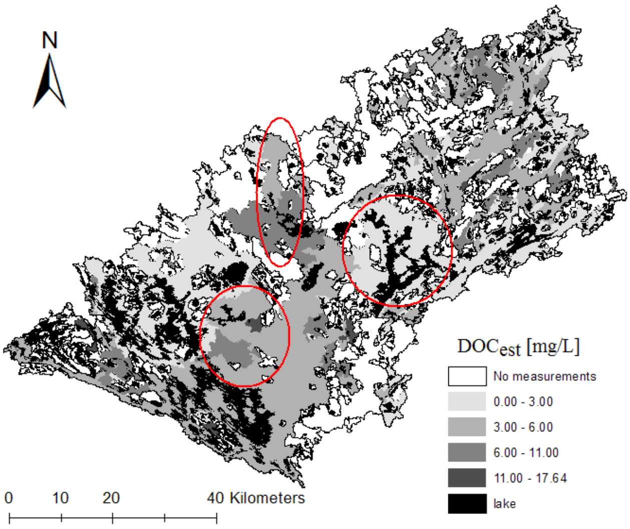


Figure 13. Estimated values for the 237 lakes with measured values for the old peat model. Look especially for differences between models and between measured values in the areas marked with a circle.

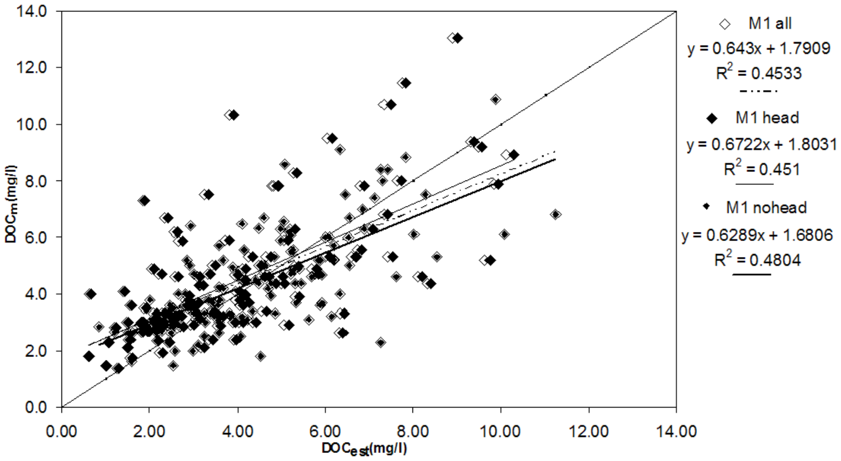


Figure 14. Model M1, headwater lakes and non headwater lakes for the optimized values of v_u and v_l. A relationship and r^2 is made for each and for the full set of 237 lakes (where measurements were available).

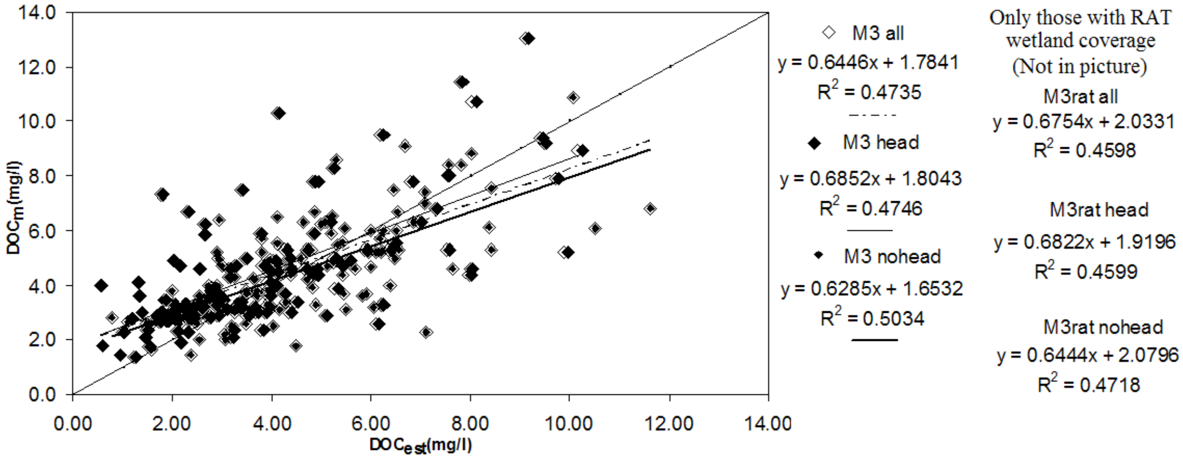


Figure 15. Model M3 headwater lakes and non headwater lakes for the optimized values of v_u and v_l. A relationship and r^2 is made for each and for the full set of 237 lakes (where measurements were available). The relationships for the same dataset containing only those with RAT wetland coverage are also shown.

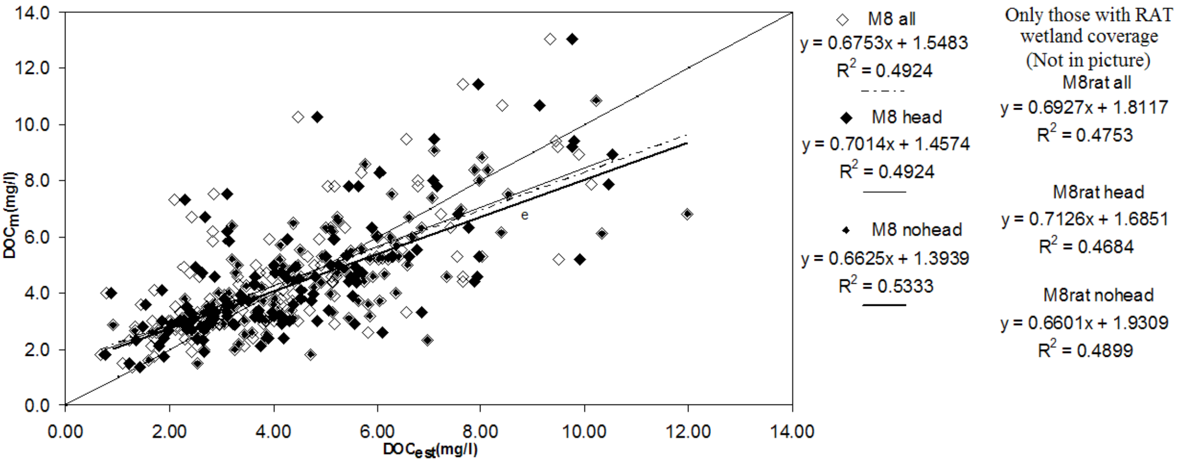


Figure 16. Model M8, headwater lakes and non headwater lakes for the optimized values of v_u and v_l. A relationship and r^2 is made for each and for the full set of 237 lakes (where measurements were available). The relationships for the same dataset containing only those with RAT wetland coverage are also shown.

7 DISCUSSION

The three models obtained do not explain as much of the variance of DOC in the 20 subcatchments of Dorset as the original model of Dillon and Molot (1997b). The correlations between the older peat percentage values from 1997 and the longer DOC series is still higher at 0.75 (the shorter series used in 1997 gave a correlation of 0.88) than for the best parameter found with GIS. This is to be compared to the lower values for percentage RAT wetlands of 0.50 and average slope of - 0.63. Grouping the new parameter datasets, as well as the peat percentage values attained with air photos and field work, with the mean DOC values for the Dorset study area showed that the peat percentage was also the only parameter grouped together with the DOC and DOC/Q.

The 1997 peat model (using the newer RAT wetland data) also explained the highest percentage of DOC, when only the 90 lakes that had RAT wetland coverage and headwater lakes were used, but for all lakes it explained the lowest percentage of all models. This might mean that the wetland data are more important than seen with the GIS for the Dorset area. It also suggests that the available GIS data might be good enough for the needs at present, but that improvements are still possible. One hope is for example that LiDAR data, resulting in DEM´s with a much greater accuracy, will become more commonly available (at present the price is an obstacle). This is particularly so since average slope of catchments seems more important than the percentage of wetlands from GIS data.

One thing noticed is that during optimization of the v_u and v_l too many different parameters are affected. It is the absolute average deviation, the average of the absolute deviations, r^2, how the result then fit to a y = x line and so on. To find the fit for the intended goals different values for the two parameters should be use to see which set fit the data best. If measured data is not available different values can be used to see how the results differ for each case to get a range of results rather than just one value.

As results were obtained for the Muskoka River Watershed before and after optimization of the loss coefficients in the Lake DOC Model, a bias was found in the residuals. With the optimized values a better distribution between positive and negative residuals was obtained, but a negative trend bias was still evident. The reason for this bias could not be determined, but it was found for all cases and all models. This gives one reason to suspect that the error lies in the Lake DOC Model rather than in each of the stream DOC Models. Other causes were considered, but not evaluated, and these were a possible regional trend and areal factors. The first cause was considered as the area of the Muskoka River Watershed is large and the Dorset study area used to develop the models (all four) is situated in only one part of the larger watershed. The models might therefore not cover some more regional factors that could explain some of the DOC differences for these areas. This hypothesis might have some support in the fact that some larger areas seem to be explained quite well by all models and others areas by neither. The second cause came up as the 20 subcatchments have small to middle sized catchment areas while the watershed shows a wider range of catchment area sizes. The smaller areas might not explain the flux in the larger areas as well as in the smaller catchments.

8 CONCLUSIONS

The models explain about the same percentage of DOC, but the average estimate of all four models is the best. This gives the conclusion that the use of several models can lead to a better result than just one.

8.1 FOR FUTURE RESEARCH

The Lake DOC Model should get some more attention. Maybe more parameters need to be added or the loss coefficients need to be optimized for different areas and different parameters to get a better knowledge of their range and what default values one should use in different types of areas.

In an area with more agriculture or otherwise quite different from the area the model was developed for, the model should not be used without modification (or revalidation against measured data). In future studies data on soil type and geological layers should also be used. Hopefully these will become more easily available.

Another factor that might be something to look at in future studies is slope. Slope might be divided into areas, flat areas (0 - n degrees), and steep areas and so on, giving different areal percentages of these groups. The slope might also be coupled with elevation, to separate the flat high and flat low areas for example. (Kara Webster, 2007)

REFERENCES

Aravena, R., S. L. Schiff, S.E. Trumbore, P. J. Dillon and R. Elgood. 1992. *Evaluating dissolved inorganic carbon cycling in a forested lake watershed using carbon isotopes.* Radiocarbon **34**: 636-645.

Bishop K., C. Pettersson, B. Allard and Y. Lee. 1994. *Identification of the Riparian Sources of Aquatic Dissolved Organic Carbon.* Environment International **20**: 11-19.

Boyer E. W., G. M. Hornberger, K. E. Bencala and D. McKnight. 1996. *Overview of a simple model describing variation of dissolved organic carbon in an upland catchment.* Ecological Modelling **86**: 183-188.

Brien C. J., W. N. Venables, A. T. James and O. Mayo. 1984. *An Analysis pf Correlation Matrices: Equal Correlations*. Biometrika **71**: 545-554.

Creed, I. F., S. E. Sanford, F. D. Beall, L. A. Molot and P. J. Dillon. 2003. *Cryptic wetlands: integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes.* Hydrological Processes **17**: 3629-3648.

Demtener F., J. Drevet, J. F. Lqamarque, I. Bey, B. Eickhout, A. M. Fiore, D. Hauglustaine, L. W. Horowitz, M. Krol, U. C. Kulshrestha, M. Lawrence, C. Galy-lacaux, S. Rast, D, Shindell, D. Stevenson, T. Van Noije, C. Atyherton, N. Bell, D. Bergman, T. Butler, J. Cofala, B. Collins, R. Doherty, K. Ellingsen, J. Galloway, M. Gauss, V. Montanaro, J. F. Müller, G. Pitari, J. Rodriguez, M. Sanderson, S. Strahan, M. Schultz, K. Sudo, S. Szopa and O. Wild. 2006. *Nitrogen and sulphur deposition on regional and global scales: a multi-model evaluation.* Re-submitted to GBC (Green Building Council).

Dillon P. J., L. A. Molot and W. A. Scheider. 1991. *Phosphorus and Nitrogen Export from Forested Streams Catchments in Central Ontario.* Journal of Environmental Quality **20**: 857-864.

Dillon, P. J. and L. A. Molot. 1997a. *Dissolved organic and inorganic carbon mass balances in central Ontario lakes.* Biogeochemistry **36**: 29-42.

Dillon, P. J. and L. A. Molot. 1997b. *Effect of landscape form on the export of dissolved organic carbon, iron, and phosphorus from forested stream catchments.* Water Resources Research **33**: 2591-2600.

Dillon P. J., K. M. Somers, J Findeis and M. C. Eimers. 2003. *Coherent response of lakes in Ontario, Canada to reduction in sulphur deposition: the effects of climate on sulphate concentration.* Hydrology and Earth System Science **7**: 583-595.

Dillon, P. J. and L. A Molot. 2005. *Long-term trends in catchment export and lake retention of dissolved organic carbon, dissolved organic nitrogen, total iron, and total phosphorus: The Dorset, Ontario, study, 1978-1998.* Journal of Geophysical Research **110**: G01002, doi:10.1029/2004JG000003.

Evans H. E., P. J. Dillon and L. A. Molot. 1997. *The use of mass balance investigations in the study of the biogeochemical cycle of sulfur.* Hydrological Processes **11**: 765-782.

Findlay S., J. M. Quinn, C. W. Hickey, G. Burrell and M. Downes. 2001. *Effects of land use and riparian flowpath on delivery of dissolved organic carbon to streams.* Limnology and Oceanography **46**: 345-355.

Freund J. E. 1967. *Modern Elementary statistics.* Third edition. Prentice-Hall, Inc. US.

Fröberg M., D. Berggren, B. Bergkvist, C. Bryant and J. Mulder. 2006. *Concentration and fluxes of dissolved organic carbon (DOC) in three Norway spruce stands along a climatic gradient in Sweden.* Biogeochemistry **77**: 1-23, doi: 10.1007/s10533-004-0564-5

Futter, M. N., D. Butterfield, B. J. Cosby, P. J. Dillon, A. J. Wade, and P. G. Whitehead. 2007. *Modeling the mechanisms that control in-stream dissolved organic carbon dynamics in upland and forested catchments*. Water Resources Research **43**: W02424, doi: 10.1029/2006WR004960.

Futter M. N. 2007. *Controls on Dissolved Organic Carbon Concentrations in Surface Waters Assessed using INCA-C, The Integrated Catchments Model for Carbon.* PhD Thesis, Trent University, Peterborough, Ontario, Canada.

Gennings, C., L. A. Molot and P. J. Dillon. 2001. *Enhanced photochemical loss of organic carbon in acidic waters.* Biogeochemistry **52**: 339-354.

Gustafsson J. P., G. Jacks, M. Simonsson and I. Nilsson. 2005. *Soil and water chemistry.* Department of Soil Sciences, SLU. Sweden.

Hudson, J. J., P. J. Dillon and K. M. Somers. 2003. *Long-term patterns in dissolved organic carbon in boreal lakes: the role of incident radiation, precipitation, air temperature, southern oscillation and acid deposition.* Hydrology and Earth System Science **7**: 390-398.

Issar A. S. 2004. *Climate Changes during the Holocene and their Impact on Hydrological systems*. International Hydrology Series. University Press, Cambridge, United Kingdom.

Jonsson A, G. Algesten, A-K. Bergström, K. Bishop, S. Sobek, L.J. Tranvik and M Jansson. 2007. *Integrating aquatic carbon fluxes in a boreal catchment carbon budget.* Journal of Hydrology **334**: 141-150. doi: 10.1016/j.jhydrol.2006.10.003

Köhler S, I. Buffam, A. Jonsson and K. Bishop. 2002. *Photochemical and microbial processing of stream and soil water dissolved organic matter in a boreal forested catchment in northern Sweden.* Aquatic Sciences **64**: 269-281

Lindsjö A. 2005. *Predicting Dissolved Organic Carbon Concentrations in Swedish Boreal Streams from Map Information.* Graduate Thesis, Swedish University of Agricultural Sciences, Sweden.

Magnuson, J. J., K. E. Webster, R. A. Assel, C. J. Bowser, P. J. Dillon, J. G. Eaton, H. E. Evans, E. J. Fee, R. I. Hall, L. R. Mortsch, D. W. Schindler and F. H. Quinn. 1997. *Potential effect of Climate changes on Aquatic Systems: Laurentian Great Lakes and Precambrian Shield Region.* Hydrological processes **11**: 825-871.

Michalzik B. and K. Kalbitz, J-H. Park, S. Solinger and E. Matzner. 2001. *Fluxes and concentrations of dissolved organic carbon and nitrogen – a synthesis for temperate forests.* Biogeochemistry **52**: 173-205

Molot, L. A. and P. J. Dillon. 1996. *Storage of terrestrial carbon in boreal lake sediments and evasion to the atmosphere.* Global Biogeochemical Cycles **10**: 483-492.

Molot, L. A. and P. J. Dillon. 1997a. C*olour - mass balances and colour - dissolved organic carbon relationships in lakes and streams in central Ontario.* Canadian Journal of Fisheries and Aquatic Sciences **54**: 2789-2795.

Molot, L. A. and P. J. Dillon. 1997b. *Photolytic regulation of dissolved organic carbon in northern lakes.* Global Biogeochemical Cycles **11**: 357-365.

Molot, L. A., W. Keller, P. R. Leavitt, R. D. Robarts, M. J. Waiser, M. T. Arts, T. A. Clair, R. Pienitz, N. D. Yan, D. K. McNicol, Y. T. Prairie, P. J. Dillon, M. Macrae, R. Bello, R. N. Nordin, P. J. Curtis, J. P. Smol and M. S. V. Douglas. 2004. *Risk analysis of dissolved organic matter-mediated ultraviolet B exposure in Canadian inland waters.* Canadian Journal of Fisheries and Aquatic Sciences **61**: 2511-2521. doi: 10.1139/F04-165

Molot, L. A., J. J. Hudson, P. J. Dillon and S. A. Miller. 2005. *Effect of pH on photo-oxidation of dissolved organic carbon by hydroxyl radicals in a coloured softwater stream.* Aquatic Sciences **67**: 189-195. doi: 10:1007/s00027-005-0754-9

Moore T. R. 2003. *Dissolved organic carbon in a northern boreal landscape.* Global Biogeochemical cycles **17**: 1109. doi: 10.1029/2003GB002050

Mulholland P. J. 2003. *Aquatic Ecosystem: Interactivity of Dissolved Organic Matter. Chapter 6; Large-Scale Patterns in Dissolved Organic Carbon Concentration, Flux, and Sources.* Elsevier Science. USA.

Neff J.C. and G. P. Asner. 2001. *Dissolved Organic Carbon in Terrestrial Ecosystems; Synthesis and a Model.* Ecosystems **4**: 29-48. doi: 10:1007/s100210000058

Neter J., W. Wasserman and M. H. Kutner. 2nd edition 1989. *Applied Linear Regression Models.* Richard D IRWIN Inc. US.

O'Connor E. M. 2007. *Modeling Mercury Concentrations in Fish of the Muskoka River Watershed – A Mass Balance Approach involving Dissolved Organic Carbon and Mercury Models.* MSc Thesis, Trent University, Peterborough, Ontario, Canada.

Quinn G.P. and Keough M.J. 2006. *Experimental design and data Analysis for Biologist.* University Press, Cambridge. United Kingdom.

Reid R. A., R. Girard and A. C. Nicolls. 1987. *Morphometry and Catchment areas for the calibrated watersheds.* Data report DR 87/4. Ministry of the Environment, Ontario (MOE).

Schiff, S., R. Aravena, S. E. Trumbore and P. J. Dillon. 1990. *Dissolved Organic Carbon Cycling in Forested Watersheds: A Carbon Isotope Approach.* Water Resources Research **26**: 2949-2957.

Schiff, S. L., R. Aravena, S. E. Trumbore, M. J. Hinton, R. Elgood and P. J. Dillon. 1997. *Export of DOC from forested catchments on the Precambrian Shield of Central Ontario: Clues from ^{13}C and ^{14}C.* Biogeochemistry **36**: 43-65.

Schiff S., R. Aravena, E. Mewhinney, R. Elgood, B. Warner, P. Dillon and S. Trumbore. 1998. *Precambrian Shield wetlands: Hydrologic control of the sources and export of Dissolved organic matter.* Climatic Change **40**: 167-188.

Schnoor J. L. 1996. *Environmental modeling. Fate and transport of pollutants in water, air, and soil.* John Wiley & sons, Inc. US.

Tabachnick B.G. and L. S. Fidell. 5th edition 2007. *Using Multivariate Statistics.* Pearson Education Inc. US.

Tan K. H. 2003. *Humic Matter in Soil and the Environment. Principles and Controversies.* Marcel Dekker, INC. US.

Vidon P. G. F. and A. R. Hill. 2004. *Landscape controls on the hydrology of stream riparian zones.* Journal of Hydrology **292**: 210-228. doi: 10.1016/j.jhydrol.2004.01.005

Vinod H. D and A. Ullah. 1981. Recent advances in regression methods. Marcel Dekkers, Inc. US.

Wu, F. C., R. B. Mills, Y. R. Cai, R. D. Evans and P. J. Dillon. 2005. *Photodegradation-induced changes in dissolved organic matter in acidic waters.* Canadian Journal of Fisheries and Aquatic Sciences **62**: 1019-1027. doi: 10.1139/F05-009

8.2 INTERNETS SITES

GraphPad: GraphPad.com –Data analysis and biostatistics software and resources, 2007-10-26, http://www.graphpad.com/library/BiostatsSpecial/article_197.htm

HT: Hawth´s Analysis Tools for ArcGIS, 2007-08-22 http://www.spatialecology.com/htools/tooldesc.php

IE: Introductory econometrics, 2007-09-25; 18 pm
http://caleb.wabash.edu/econometrics/EconometricsBook/
Add-in, MCSim: http://caleb.wabash.edu/econometrics/EconometricsBook/Basic%20Tools/ExcelAddIns/MCSim.htm

MWCI: Muskoka Watershed Council Information, 2007-07-16, 5:02 pm http://www.muskokaheritage.org/watershed/watershedinformation.asp

OBM: The geographic networks, Ontario Basic Mapping, 2007-07-31 http://www.geographynetwork.ca/website/OBM/

SERC: Smithsonian Environmental research Center, 2008-01-20, 6:00 pm http://www.serc.si.edu/labs/co2/c3_c4_plants.jsp

8.3 ORAL REFERENCES

Aherne, Julien. Trent Univeristy, Department of Environmental & Resource Studies, Peterborough

Dillon, Peter. Supervisor, Trent Univeristy, Department of Environmental Science Peterborough.

Findeis, Joseph. Researcher on Dorset Study project, Dorset.

Webster K. nov 2007. *Topographic Controls on Carbon Dioxide Efflux from Forest Soils.* Unpublished PhD Thesis presentation. University of Western Ontario.

APPENDIX A – THE SUBCATCHMENTS OF THE SEVEN LAKES IN THE DORSET STUDY

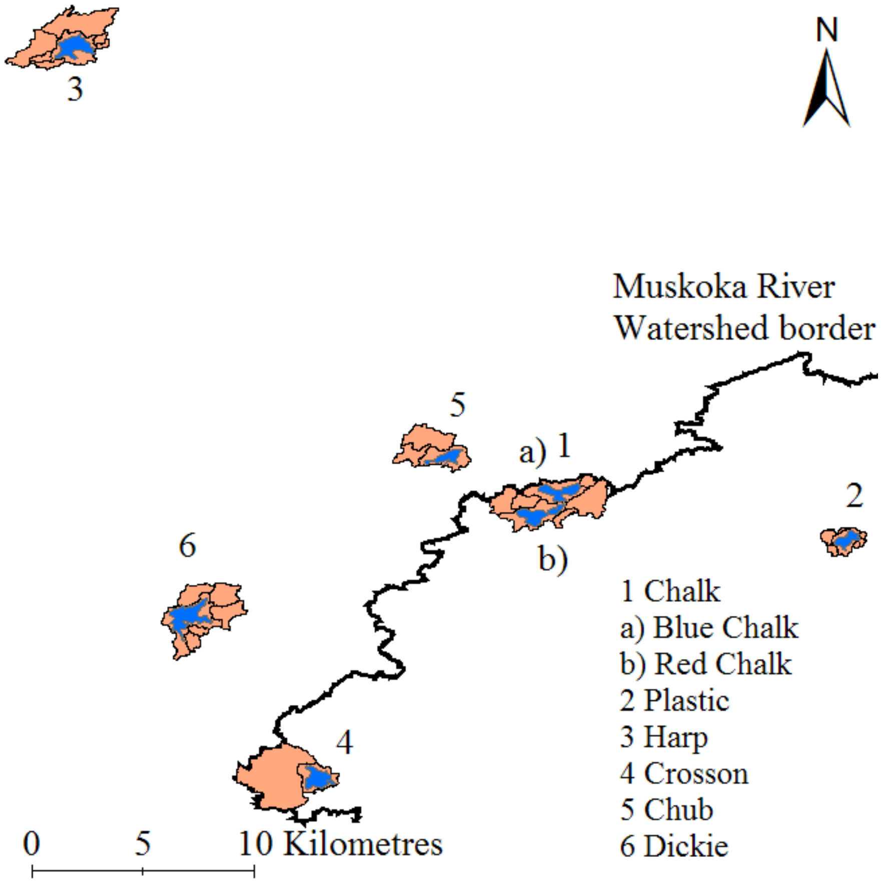


Figure App- 1. The seven lakes used for development of the model and their 20 subcatchments. The black line is the border to the Muskoka river watershed (where the model was later applied on catchments). As can be seen three lakes are inside this area and five are outside (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

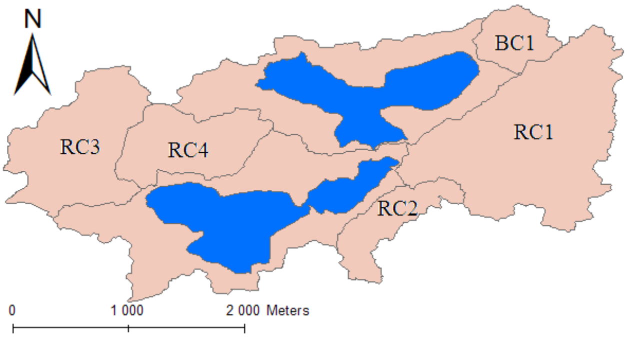


Figure App- 2. Red Chalk and Blue Chalk lakes, the lake to the north being Blue Chalk (Red being downstream of Blue Chalk), and their subcatchments. Only one subcatchment belongs to Blue chalk (BC1) and four to Red Chalk (RC1, RC2, RC3 and RC4) (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

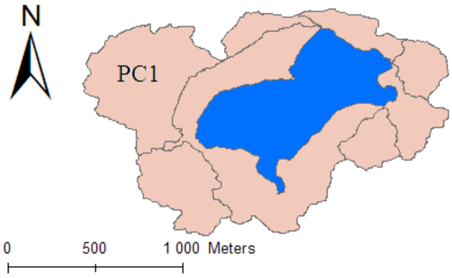


Figure App- 3. Plastic Lake and its six subcatchment, of which only PC1 have data of stream DOC available (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

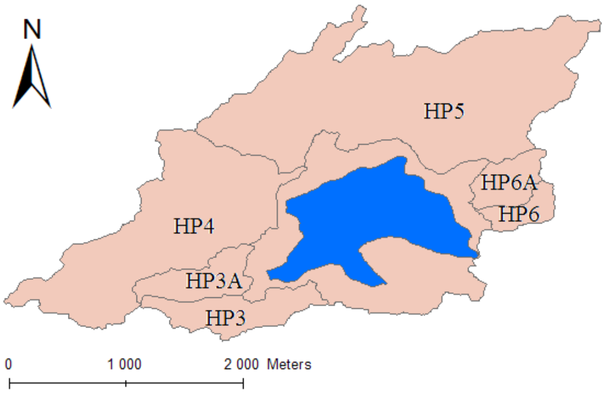


Figure App- 4. Harp Lake and the six subcatchments (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

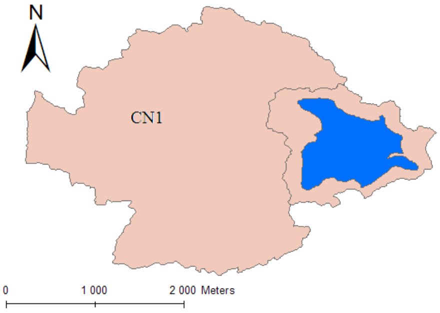


Figure App- 5. Crosson lake and CN1 subcatchment, taking up most of the total catchment area (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

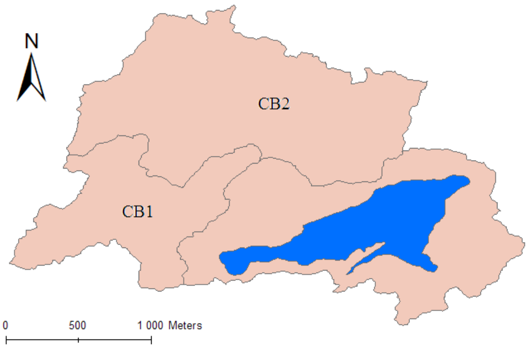


Figure App- 6. Chub Lake and the two subcatchments, CB1 and CB2 (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

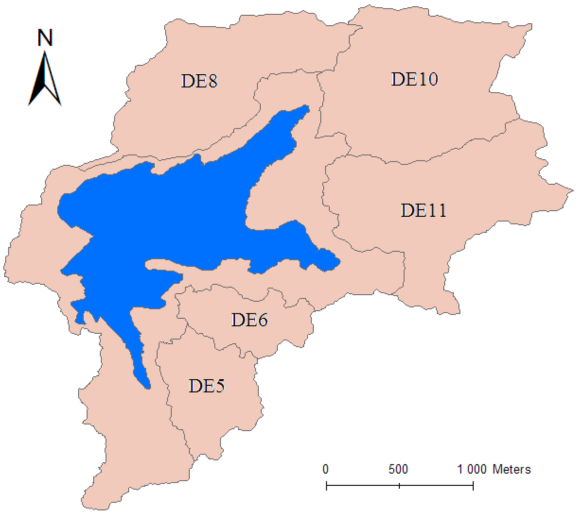


Figure App- 7. Dickie Lake and the five subcatchments; DE5, DE6, DE8, DE10 AND DE11 (Reid *et al.,* 1987). (Picture made in ArcGIS 9.1.)

Table App- 1. Data of Mean annual DOC and Percent of peat used to attain the original mass balance model by Dillon and Molot (1997b) as well as the percentage of minor till plain, thin till and ponds in the subcatchments. The last column also shows the number of years with DOC measurements available between 1978 and 1998 in each subcatchment, the data that is used to attain the new mass balance model. X1 is average slope, X2 is percentage wetlands and X3 is drainage density.

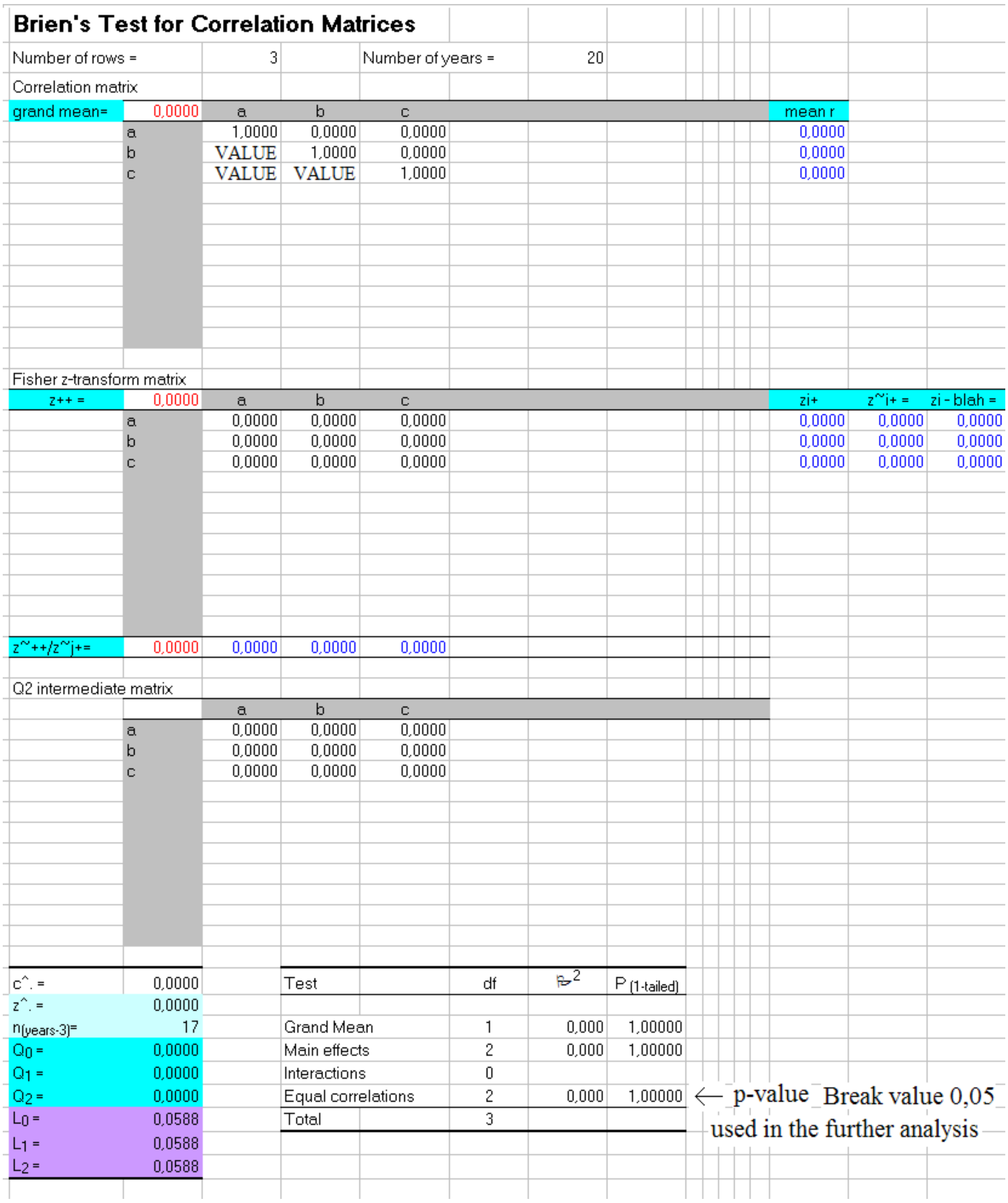
	Sub		Catchment	Minor till	Thin			Number of measure-	DOC	Q	X1	X2	X3
Lake	catchment	DOC	Area	Plain	Till	Peat	Ponds*	ments of DOC					
		m^2/y	ha	%	%	%	%	Number of years	mg/l	m^3	º	%	m^{-1}
Blue Chalk	BC1	990	20.4	94	6	0	0	16	1031	0.407	9.13	0	0.0E+00
Chub	CB1	2290	59.7	24.2	72.4	2.8	0.6	20	2552	0.422	5.65	5.19	1.7E-03
	CB2	6020	126	16.7	75.3	8	0	20	6713	0.526	5.03	4.16	1.9E-03
Crosson	CN1	4360	456.3	17.1	67.1	8.6	7.2	12	4346	0.547	3.72	12.28	1.4E-03
Dickie	DE11	6570	78.9	0	82.9	17.1	0	17	6819	0.535	3.69	15.16	1.1E-03
	DE10	8550	76.3	0	79.1	20.9	0	20	9362	0.542	2.15	15.12	4.8E-04
	DE5	7310	30	0	74.6	25.4	0	20	7333	0.564	2.91	37.69	0.0E+00
	DE6	9080	21.8	0	78	22	0	20	9221	0.57	4.25	17.51	0.0E+00
	DE8	6810	67	13.7	78.1	8.2	0	20	6866	0.565	4.12	10.76	1.9E-03
Harp	HP3	4560	26	79.5	11.2	9.3	0	20	4695	0.612	8.6	6.94	2.0E-03
	HP3A	1930	19.7	97.1	0	2.9	0	20	1992	0.592	10.83	0	1.9E-03
	HP4	2990	119.5	56.1	32.8	0	11.1	20	3264	0.56	7.88	12.52	3.1E-03
	HP5	5580	190.5	34.5	48.6	13.3	3.6	20	6005	0.618	7.29	19.79	1.9E-03
	HP6	3280	10	45.2	54.8	0	0	20	3588	0.613	11.02	0	4.5E-03
	HP6A	3270	15.3	6.6	84.9	8.5	0	20	3561	0.498	7.06	0	4.8E-03
Plastic	PC1	4860	23.3	9.6	80.2	7	3.2	19	5149	0.555	4.59	12.15	9.4E-04
Red Chalk	RC1	1900	133.6	53.2	41.1	0	5.7	20	2124	0.534	6.99	2.95	2.4E-03
	RC2	6220	27	0	67.9	10.5	21.6	20	6713	0.535	5.55	0	2.7E-04
	RC3	4170	70.5	81.7	2.7	9.9	5.7	20	4561	0.63	9	16.26	1.6E-03
	RC4	3470	45.5	76.3	16	2.9	4.8	20	3884	0.571	10.01	4.77	3.5E-04
					Sum of measurements:			384					

Source: Dillon and Molot (1997b)

* This is given as the difference between 100 % and the sum of the three other columns to the left.

APPENDIX B – BRIEN´S TEST EXCEL WORK SHEETS.

The Excel spreadsheet in the Figure App-8 was developed by Brien *et al.* (1984) as it uses the method developed in this work. This spreadsheet is where the computations are made to determine if the last parameter entered into the matrix on the top of the sheet (in matrix correlations are entered at VALUE) should be entered into a group with the others. The choice of which to enter is based on the highest correlations and the highest average of correlations for the parameters in the group already. (The working order was attained by personal communication with J. Findeis.)



Brien's Test for Correlation Matrices

Number of rows = 3 | Number of years = 20

Correlation matrix

grand mean=	0,0000	a	b	c	mean r
	a	1,0000	0,0000	0,0000	0,0000
	b	VALUE	1,0000	0,0000	0,0000
	c	VALUE	VALUE	1,0000	0,0000

Fisher z-transform matrix

z++ =	0,0000	a	b	c	zi+	z~i+ =	zi - blah =
	a	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	b	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	c	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
z~++/z~j+=	0,0000	0,0000	0,0000	0,0000			

Q2 intermediate matrix

		a	b	c
	a	0,0000	0,0000	0,0000
	b	0,0000	0,0000	0,0000
	c	0,0000	0,0000	0,0000

c^. =	0,0000
z^. =	0,0000
$n_{(years-3)}$=	17
Q_0 =	0,0000
Q_1 =	0,0000
Q_2 =	0,0000
L_0 =	0,0588
L_1 =	0,0588
L_2 =	0,0588

Test	df	χ^2	P (1-tailed)
Grand Mean	1	0,000	1,00000
Main effects	2	0,000	1,00000
Interactions	0		
Equal correlations	2	0,000	1,00000
Total	3		

← p-value Break value 0,05 used in the further analysis

Figure App- 8. Brien´s Test Excel sheet for three inputs and the number of data points, or years or data 20.

APPENDIX C – COMPUTER PROGRAMS USED

The programs used during different parts of this project are:

- ArcGIS 9.1, student edition from ESRI (Environmental Systems Research Institute). The parts used: ArcMAP and ArcCatalog. VBA codes and extensions tools (HawthsTools, ArcToolbox, Editor and Spatial Analyst Tools mostly) were used.
- Microsoft ® (XP Home edition):
  - Access – for Excel files that would be imported into ArcGIS 9.1.
  - Excel - for compiling of data, regression and analysis (most other statistical programs used had the result exported to Excel). VBA programming was also used inside this platform. Excel is used if no other statistical software is specifically mentioned.
  - Paint - where pictures were made or altered.
  - PowerPoint - presentation
  - Word - the report.
- Statistical software, outside of Excel:
  - SPSS – evaluation version of 15.0. Used for regression.
  - S-PLUS – 8.0, student version. Used for regression.
  - GraphPads Prism 5 – evaluation version. Used for normality tests.
  - Kyplot (histograms), Jump and Minitab 15 were mainly used to compare results between programs to investigate if they gave similar regression coefficients. There were some differences, but as SPSS and S-PLUS gave the most similar results and where user friendly they were mostly used (S-PLUS, mostly as the evaluation period for SPSS expired prior to the end of the thesis project).
  - Crystal Ball 7, evaluation version - used for optimization of the Lake DOC Model.

APPENDIX D – ACCURACY OF GIS LAYERS

Each layer in GIS has a specific accuracy which depends on the data behind the layer, where it came from and how accurate that method was. For most layers there is a horizontal accuracy only, but for layers like DEM´s there is also a vertical accuracy. The accuracy of these layers affects the accuracy of layers coming from them, like slope or distance from lake and so on. Usually the accuracy of a layer can be found in the metadata – a text document that is to be sent along with the layer files from the source. For the road data no metadata could be obtained.

- **DEM**
    - Horizontal accuracy: Precise: ± 10 m
    - Vertical accuracy: Reliable: ± 5 m
- **Forest (FRI)**
    - Horizontal absolute accuracy: always within 10 meters
    - Horizontal relative accuracy: within 2.5 meters
    - Area: within 0.10 ha
- **Streams (flow accumulation data) and catchments**
    - Came as a result of delineation of older DEM´s. Their accuracy was not known, but as delineation was performed in the early part of the century, the accuracy of DEM´s from this is assumed to be ten m.
- **Lakes, and ponds**
    - Derived by Perry (2001)
- **RAT (Ducks unlimited) Wetlands**
    - Different flight tests have been made to ascertain the accuracy of the RAT technique to estimate wetlands area in the region. One in Haliburton (just outside of the area, to the east) 2004 and one in part of Parry Sound (2005). The first found 90 % to be correctly identified and a positional accuracy: ca ± 35 m. The second test flight found that some segments were missing in some wetlands, giving an underestimation and that the classification of bogs and fens needed improvement. (O´Connor, 2007)
- **NRVIS (MNR) wetlands and ponds**
    - Horizontal accuracy: Reliable: ± 100 m

APPENDIX E – WAYS OF ATTAINING PARAMETERS FROM GIS

Dorset study area:

All parameters below were calculated for each catchment with the help of *ArcMAP*. All layers had (or was altered to) the projection *NAD_83 UTM_17N*. UTM stands for Universal Transverse Mercator and it has 60 zones covering the world. Each of these stretches six degrees in longitude and has the highest accuracy at the centre. All layers were also cut (with *clip* for features and *extract by mask* for rasters, both in the extension *ArcToolbox*) with the polygons for each of the 20 subcatchments. This meant that areas and length would need to be recalculated and this was done in the fields in the *attribute Tables* for each layers with the following VBA code:

```
Area:   Dim dblArea as double          Length:   Dim dblLength as double
        Dim pArea as Iarea                       Dim pCurve as ICurve
        Set pArea = [shape]                      Set pCurve = [shape]
        dblArea = pArea.area                     dblLength = pCurve.Length
```

Computations of percentage, quotas and so on were made once values had been added to Excel.

- □ Average catchment slope → one parameter

Slope was calculated from DEM´s. The DEM´s and slope layers calculated from them were both, for comparison, merged in *ArcMAP*. (The average slope reached for both cases seemed identical.) As both the DEM´s and the slope layers are rasters they where extracted by: "Extraction by mask", in *Arc Toolbox* > *Spatial Analyst Tools* > *Extraction* > *Extract by Mask.* The average slope was attained from raster statistics and the separate means were entered into Excel. The parameter was called → **avslope** (the list of parameters and their abbreviations can be found in list of Abbreviation).

- □ Catchment area, catchment perimeter and quota between them → three parameters

Area and perimeter for the catchments could be obtained from fields in the Attributed Tables of the catchment polygons. The quota area/perimeter was then computed in Excel. The parameters were called, in order → **catarea**, **catperi** and **area_peri**

- □ lake area/catchment area → one parameter

Smaller lakes (< 5 ha) had first to be removed from the layer (these where later used in a pond layer called spond, and as wetlands together with both wetland types in perRAT2 and perOBM2 (see also below)). The area of lakes could also be obtained from the polygons for lakes and the quota was computed in Excel. The parameter was called → **arealake_cat**

- □ Wetland percentage – from two sources → two parameters (+ 4 under ponds, as the layers were merged and seen as wetlands)

Area was recalculated and then entered into Excel where the percent was computed.

The wetland layer from MNR/NRVIS, contained wetlands permanent and water bodies. Wetlands permanent were used (small water bodies < 5 ha was seen as ponds, wpond). The parameter was called → **perOBM** (percentage NRVIS wetlands). Wetlands were present in the following six subcatchments: HP4, HP5, CB2, DE8, DE10 and PC1.

The wetland layer from Ducks unlimited gave the parameter that was called → **perRAT** (percentage RAT wetlands) as the data was attained with the Rapid Assessment Technique (O´Connor, 2007). Data was present in 15/20 subcatchment, namely: HP3, HP4, HP5, CB1, CB2, RC1, RC3, and RC4, all at Dickie Lake, CN1 and PC1.

- ❑ Small lakes/ponds – from two sources, used separately and together with wetlands → six parameters

The area of ponds was recalculated and for the cases where ponds were merged with the wetlands another recalculation occurred after the layers were dissolved (to avoid counting overlaps twice). Dissolution: *Arc Toolbox > Data Management Tools > Generalization > Dissolve*, was made with *single part*).

Ponds from Ducks unlimited were, as mentioned small lakes (< 5 ha) from the lake layer and the parameters were called → **perRAT2** and **perOBM2**, the pond themselves **spond**. Ponds were present only in: CN1 and RC1.

Ponds also came from the water layer that gave NRVIS wetlands, the part, named as water bodies (those < 5 ha) and the parameters were called → **perRAT3** and **perOBM3**, the ponds themselves **wpond**. Ponds were present in eight subcatchments: HP4, HP5, CB2, CN1 and all subcatchments for lake Red Chalk.

- ❑ Forest percentage – on catchment and wetland (not with ponds) → three parameters

Recalculated areas were imported to Excel, where percent was computed.

The forest layers came from the local MNR offices FRI (Forest Resource Inventory) departments (Parry Sound and Bancroft) as well as the Bata Library (Algonquin Park). The layers were merged and then only the polygons with the value FOR (which stands for forest according to the MNR metadata) in the field POLYTYPE were selected. The parameter was called → **perFOR** (percentage forest on catchment area). Forest was present in all subcatchments.

The forest layer was also cut for each subcatchment wetland layers (RAT and OBM wetlands) to gain percentages of the wetlands covered by forest → **perFORRAT** and **perFOROBM** (as in percentage of RAT/OBM wetland covered with forest).

- ❑ Straight line distance between lake and wetland (both sources) – average, max and min distance → six parameters

First a straight line distance layer (*Spatial Analyst Tools*) was computed and then this was cut for each subcatchment. As not all streams were measured at the inflows to the lakes, but somewhat upstream, this might give an error. The error is demeaned minor though as most measurement points were within 20 meters of the lake inlet (personal communication, Peter Dillon). The coordinates for the measurement points were also not so accurate (20-30 meter errors, personal communication, Peter Dillon and Joe Findeis) to make it worth the effort of computing distance from them → the basic outline of the parameters were **distlake(RAT/OBM)av/max/min**, for example **distlakeRATav**, meaning average distance to lake from RAT wetlands

- ❑ Percentage of road length to catchment area → one parameter

The road layers came from the geographic network and were downloaded from their internet site (www, OBM, 2007). The data obtained was merged, dissolved and cut for the subcatchments before the road length was recalculated and the sum entered into Excel

The parameter was called → **perRoad**. Data over roads, were not present in the following four subcatchments: HP6A, HP3A, DE10 and RC2.

- ❑ Drainage density = stream length/catchment area, Stream average slope and stream average slope/stream length → three parameter

More than one stream layer was obtained, but most of these had very few actual streams, only data from Ducks unlimited had data considered good enough. Even so three of the 20 subcatchments (BC1, DE5 and DE6) did not have streams in this layer, even if in reality the streams have been measured for DOC. This layer was converted from raster to feature (polyline shape file), both layers were cut for each subcatchment. The stream length was recomputed in the feature layer and in Excel Drainage density was then calculated as stream length/catchment area. Average slope was obtained by extracting the slope layer with the stream (raster) layer for each subcatchment. After input into Excel the third parameter was also computed. The parameters were then called in order → **drainden**, **strslope** and **strslope_len**

- ❑ Agriculture turned out to be close to non-existing in the region, not surprising given that the area lies on the Precambrian shield and has a very thin soil cover. It is though advised if this model is to be used in an area with more agriculture to look into redoing the regression with agriculture as a factor.
- ❑ Soil layers were not available for this area at this time.
- ❑ Open land was not available at this time.
- ❑ Bedrock, came from two sources:

As the POLYTYPE RCK – rock in the FRI data, but it was sent only for one of the FRI layers and only one of the subcatchments had a polygon of this type (more catchments were covered by the layer). It was therefore decided that it would not be used.

From the Bata Library, where different types of bedrock were available, but only two types of bedrock were found in the 20 subcatchments and all but one (PC1, second type below) had the same. It was determined that the dataset was too small and similar to be able to draw any conclusions from it. The two bedrock types were defined as:

- Commonly layered biotites gneisses and migmatites; locally includes quartzofeldspatic gneisses, ortogneisses, paragneisses
- tonalite, grandodiorite, monzonite, granite, syneite, derived gneisses

With the available layers a total of **26** parameters were obtained, analysed and used to gain models for DOC.

Muskoka River Watershed:

Three parameters were needed to be found for the Muskoka River Watershed. Three different models were run for the area, with 1, two and three parameters in a natural series. First the necessary layers were merged for the three areas (for the DEM´s more layers needed to be merged), dissolved (for the wetland and stream layers due to overlapping) and cut to cover the whole of the MRW. A new catchment layer was formed, that did not contain lakes so that the parameters would be computed only for land area. The parameter was then calculated for each catchment polygon in the watershed with an extension called *HawthsTools* (www, HT, 2007). This needed to be done for mainland and islands separately as the joint layer had irreparable geometry errors. The results from the two layers for each catchment were then put together in Excel.

- Average slope, **avslope**:

HawthsTools > Raster analysis > Zonal statistics (++)

This gave average, min, max and sum in a separate table that was then joined to the lake MasterID and exported as a .dbf4 file.

- Percentage RAT wetland, **perRAT**:

Two different layers needed to be attained to gain the RAT wetlands, as there was a gap in data in the north of the region (see Figure App-9). The other kind of wetlands, NRVIS, did not have this gap. When also attaining the percentage of NRVIS wetlands a linear relationship could be obtained from the 756 catchments in the region covered with both wetland layers. This was then used to fill up the gap in RAT wetlands.

First *HawthsTools* was run for both the wetland layers to gain the areal coverage for each catchment:

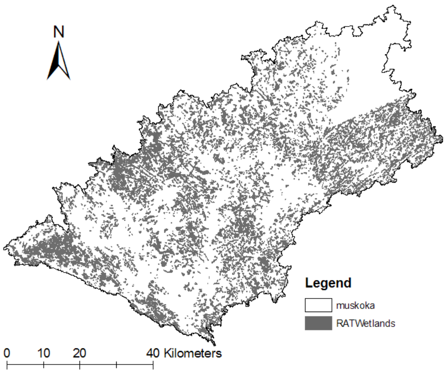


Figure App- 9. The gap in the Ducks unlimited wetland layer is evident in the northern part of the region. (Picture made in ArcMAP.)

HawthsTools > Analysis tool > polygon in polygon analysis

"The Area based summary ..." was used in this tool and this summaries the area that is within the boundaries of each catchment polygon. The result was joined with the catchment layer to gain the MasterID and exported to Excel. There the percentage was computed for the mainland and islands together for each catchment. To fill up the gap in the wetland data different relationship between the wetland types where tried. The one used then filled up the gaps as well as the other wetland type directly and both of them were used for different model spreadsheets.

- Drainage density, **drainden**

It was obtained by first getting the sum of stream length for each catchment polygon with: *HawthsTools > Analysis tool > Sum Line Lengths in Polygons*

This was exported and drainage density was computed in Excel as summed stream length/catchment area.

The function *VLOOKUP* was then used for each model spreadsheet to locate the data of the pertinent parameters (x_1, x_2 and x_3) as well as catchment and lake area (already present but a comparison was made). The data from ArcGIS were added to another spreadsheet, from which they could be located.

APPENDIX F - LIST OF COLUMNS IN THE LAKE DOC MODELS EXCEL SHEET

Column number	Name of column	unit	Explanation
1	ID	-	
2	x-coordinate	UTM	
3	y-coordinate	UTM	
4	MasterID	-	Used to identify X1, X2 and X3 and to place them right in the model spreadsheet
5	DOWNSTRMID	-	
6	NUMLAKESCONTRIB	-	Number of upstream lakes
7	X1	degrees	Average slope used in all models:
8	Lake area	m^2	From GIS
9	Wetland area	m^2	Not used
10	Upland area	m^2	Not used
11	Catchment – lake area	m^2	From GIS
12	Catchment area	m^2	C8+C11 (C stands for column, 8 and 11 are column numbers)
13	X2	%	Percentage RAT/Ducks unlimited wetlands, Model 3 and 8
14	X3	/m	Drainage density, Model 8
15	Model equation (one per time)	$mg/m^2/yr$	Equations of models, M1, M3 or M8. Gives DOC
16	DOC stream input from catchment	mg/yr	C15*C11
17	DOC load from stream export	$mg/m^2/yr$	C16/C8
18	DOC load from upstream lakes	$mg/m^2/yr$	C25, from all upstream lakes, divided by lake area (for this catchment)
19	Water from upstream lakes	$mg/m^2/yr$	sum of columns C21*C8 for all upstream lakes
20	Direct catchment runoff	m/yr	0.001*C29
21	Lake discharge, q	m/yr	(C19+C20*C11)/C8
22	Lake DOC	mg/l	$0.001*(C18/(C21+v_l)+C17/(C21+v_u)$ (DOC_{est})
23	Measured DOC	mg/l	Observed/measured value of DOC (DOC_m)
24	Total TOC load to lake	mg/yr	C16+C17*C8
25	DOC discharge from lake	mg/yr	C8*C21*C21*1000
26	POC sediment storage	mg/yr	0.68*(C24-C25)
27	CO^2 evasion	mg/yr	0.32*(C24-C25)
28	Lake names	-	
29	Map data Runoff	mm/yr	
30	Residuals	mg/l	C23-C22
31	Absolute value of residual	mg/l	Abs(C30)
32	RAT wetlands	-	No if outside of layer

APPENDIX G – CORRELATIONS, REGRESSIONS AND GROUPINGS OF PARAMETERS FROM THE DORSET STUDY.

Table App- 2. Correlations matrix with the 26 parameters and DOC as well as DOC/Q. Numbers of parameters can be found in the List of Abbreviations.

	DOC	DQ*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
DQ	0.9																										
1	-0.63	-0.75																									
2	0.05	0.05	-0.27																								
3	-0	-0.01	-0.24	0.94																							
4	0.02	0.02	-0.28	0.93	0.94																						
5	0.15	0.17	-0.31	-0.03	-0.05	0.01																					
6	0.5	0.5	-0.55	0.21	0.16	0.2	0.07																				
7	-0.21	-0.26	0.43	-0.23	-0.17	-0.25	-0.02	-0.15																			
8	0.13	0.14	-0.28	-0.12	-0.11	-0.08	0.96	0.06	-0.01																		
9	0.37	0.42	-0.46	0.2	0.04	0.12	0.43	0.37	-0.11	0.38																	
10	0.19	0.21	-0.18	0.47	0.32	0.44	0.29	0.16	-0.03	0.06	0.24																
11	0.15	0.16	-0.17	0.43	0.29	0.41	0.41	0.15	-0.02	0.17	0.31	0.97															
12	0.18	0.19	-0.15	0.49	0.35	0.45	0.18	0.21	-0.02	-0.02	0.22	0.97	0.92														
13	0.19	0.22	-0.35	0.72	0.57	0.67	0.02	0.3	-0.2	-0.15	0.45	0.58	0.56	0.54													
14	0.18	0.19	-0.35	0.85	0.73	0.79	0.01	0.43	-0.2	-0.14	0.41	0.59	0.59	0.6	0.92												
15	0.17	0.23	-0.24	0.29	0.13	0.2	0.01	0.03	-0.07	-0.12	0.39	0.35	0.3	0.24	0.8	0.52											
16	-0.19	-0.2	-0.14	0.63	0.78	0.7	-0.13	-0.06	-0.24	-0.1	-0.18	-0.17	-0.18	-0.16	0.19	0.33	-0.11										
17	-0.22	-0.27	0.21	0.4	0.45	0.5	-0.24	-0.12	-0.37	-0.2	-0.15	-0.23	-0.24	-0.22	0.16	0.17	-0.04	0.63									
18	0.47	0.46	-0.56	0.31	0.28	0.31	0.05	0.99	-0.18	0.05	0.34	0.15	0.14	0.21	0.33	0.48	0.01	0.08	-0.03								
19	0.45	0.44	-0.49	0.29	0.26	0.31	0	0.96	-0.24	0.01	0.32	0.08	0.07	0.14	0.32	0.45	0.02	0.09	0.15	0.97							
20	0.02	0.03	-0.36	0.33	0.4	0.41	0.82	0.03	-0.15	0.81	0.28	0.16	0.27	0.07	0.13	0.2	-0.05	0.45	0.14	0.09	0.06						
21	-0.09	-0.12	-0.04	0.33	0.36	0.45	0.5	-0.05	-0.34	0.51	0.17	-0	0.08	-0.07	0.15	0.16	-0.03	0.47	0.72	0.01	0.13	0.72					
22	-0.03	-0.07	0.34	-0.71	-0.55	-0.7	-0.22	-0.19	0.35	-0.12	-0.43	-0.45	-0.45	-0.43	-0.73	-0.72	-0.47	-0.28	-0.4	-0.24	-0.29	-0.36	-0.51				
23	0.06	0.03	-0.06	-0.24	-0.23	-0.22	0.45	0.22	-0.03	0.51	0.11	-0.04	-0.01	-0.07	-0.22	-0.16	-0.16	-0.15	-0.32	0.19	0.13	0.32	0.03	0.22			
24	-0.38	-0.42	0.42	0.01	0.03	-0.03	-0.01	-0.42	0.12	-0.06	-0.34	0.11	0.15	0.1	-0.06	-0.03	-0.09	0.06	-0.1	-0.4	-0.46	0.02	-0.1	0.35	0.01		
25	-0.26	-0.3	0.46	-0.12	-0.16	-0.15	0.07	-0.45	0.18	0.02	0.06	0.04	0.1	0.02	0.05	0	0.13	-0.16	0	-0.47	-0.45	-0.03	0.05	0.16	-0.01	0.59	
26	-0.02	-0.03	0.29	-0.37	-0.34	-0.38	0.02	-0.38	0.02	0.09	0.02	-0.3	-0.27	-0.3	-0.39	-0.44	-0.16	-0.22	0.22	-0.41	-0.29	-0.1	0.21	0.18	-0.12	-0.14	0.48

* DQ, stands for DOC/Q

Table App- 3. Groups of parameters, obtained from Brien´s test. Parameters are for the 20 subcatchments from GIS as well as the old parameters used by Dillon and Molot (1997b).

Groups	1	2	3	4	5	6	7	8	9
Members	perRAT perRAT2	distlakeOBMav distlakeOBMmin	perOBM perFOROBM	Catperi Catarea Area_peri	DOC DOC_Q Peat_PD	distlakeRATav distlakeRATmaxn	Spond Wpond perOBM3	Q, Avslope perFOR, perOBM2 distlakeOBMmax arealake_cat perRoad, drainden strslope, strslope_len	perFORRAT distlakeRATmin perRAT3

APPENDIX H – NORMALITY TESTS AND RANGE OF PARAMETERS

Table App- 4. Datasets statistics, minimum, maximum, median, mean and 25 and 75% percentiles as well as standard deviation and standard error of the dataset.

Dataset	Area	Minimum	25% Percentile	Median	75% Percentile	Maximum	Mean	Standard Deviation	Standard Error
DOC$_m$	MRW	1.40	3.100	4.000	5.300	13.10	4.534	2.005	0.130
avslope	MRW	0.826	4.204	5.668	7.058	15.74	5.757	2.067	0.0705
perRAT	MRW	0.0	1.860	7.080	13.51	93.40	9.648	10.31	0.352
perOBM	MRW	0.0	0.6564	2.980	6.560	54.52	5.083	6.671	0.228
ratOBM	MRW	0.0	3.348	7.581	13.86	93.40	10.15	9.989	0.341
RATlintrend	MRW	0.0	4.571	8.856	14.24	93.40	10.87	9.739	0.332
drainden	MRW	0.0	6.2E-4	1.3E-3	1.9E-3	5.7E-3	1.3E-3	8.8E-3	3.0E-5
streamlength	MRW	0.0	463.4	1.7E+3	5.1E+3	6.8E+5	8.4E+3	3.6E+4	1.2E+3
catarea	MRW	8.3E+4	6.1E+5	1.3E+6	3.2E+6	4.0E+8	5.0E+6	2.1E+7	7.2E+5
lakearea	MRW	5.0E+4	7.4E+4	1.2E+5	2.6E+5	1.2E+8	8.1E+5	6.5E+6	2.2E+5
Q	Dorset	0.2131	0.4591	0.5384	0.6426	0.895	0.552	0.1232	6.3E-3
DOC	Dorset	641.8	3.0E+3	4.7E+3	6.7E+3	1.4E+4	5.0E+3	2.6E+3	132.5
avslope	Dorset	2.150	4.153	6.320	8.900	11.02	6.474	2.692	0.602
catarea	Dorset	9.9E+4	2.2E+5	5.2E+5	1.1E+6	4.6E+6	8.1E+5	1.1E+6	2.3E+5
catperi	Dorset	1642	2.5E+3	3.8E+3	6.1E+3	1.2E+4	4.5E+3	2.7E+3	602.6
area_peri	Dorset	60.06	84.95	125.0	183.1	365.9	141.6	70.87	15.85
obm	Dorset	0.0	0.0	0.0	6.9E+3	5.0E+4	6.5E+3	1.3E+4	2.9E+3
rat	Dorset	0.0	4.5E+3	3.9E+4	1.2E+5	5.6E+5	9.3E+4	1.4E+5	3.1E+4
for	Dorset	2.6E+4	2.0E+5	3.3E+5	7.5E+5	3.E+6	6.6E+5	8.7E+5	1.9E+5
forobm	Dorset	0.0	0.0	0.0	0.0	5.0E+4	3.4E+3	1.2E+4	2.6E+3
forrat	Dorset	0.0	1.7E+3	2.8E+4	5.9E+4	2.7E+5	4.7E+4	6.4E+4	1.4E+4
obmav	Dorset	0.0	0.0	0.0	242.0	949.7	167.8	301.9	67.52
obmmax	Dorset	0.0	0.0	0.0	467.0	1.1E+3	222.7	381.3	85.27
obmmin	Dorset	0.0	0.0	0.0	130.4	722.0	124.7	234.0	52.31
ratav	Dorset	0.0	242.5	422.8	678.8	1.0E+3	404.1	297.0	68.13
ratmax	Dorset	0.0	82.50	607.8	970.0	1.5E+3	607.3	484.1	108.2
ratmin	Dorset	0.0	19.04	130.0	308.7	832.4	192.7	225.0	50.31
spond	Dorset	0.0	0.0	0.0	0.0	3.2E+5	1.9E+4	7.3E+4	1.6E+4
wpond	Dorset	0.0	0.0	0.0	1.5E+4	3.7E+5	3.0E+4	8.4E+4	1.9E+4
rat2	Dorset	0.0	1.039	8.853	15.99	37.69	10.24	9.543	2.134
rat3	Dorset	0.0	3.749	11.19	16.92	37.69	11.13	9.428	2.108
obm2	Dorset	0.0	0.0	0.0	2.088	8.063	1.585	2.750	0.615
obm3	Dorset	0.0	0.0	0.9453	7.197	8.115	2.857	3.395	0.759
lakearea	Dorset	3.2E+5	5.6E+5	7.0E+5	8.6E+5	9.2E+5	6.5E+5	2.0E+5	4.4E+4
lake/cat	Dorset	0.123	0.572	1.304	2.970	7.064	1.959	1.794	0.401
roadlength	Dorset	0.0	0.0173	0.119	0.257	0.547	0.154	0.155	0.0346
Strslope	Dorset	0.0	2.195	3.860	5.160	8.510	3.667	2.242	0.5013
streamlength	Dorset	0.0	0.00084	0.0035	0.0154	0.0380	0.00853	0.0110	0.00243
drainden	Dorset	0.0	0.00038	0.0017	0.0020	0.0048	0.0016	0.0014	0.00031

Table App- 5. T-test and normality test on the different datasets, Skewness and kurtosis is also computed.

		Kolmogorov-Smirnovs normality test		D'Agostino & Pearson omnibus normality test		Shapiro-Wilk normality test		One sample t	Wilcoxon Signed Rank	Exact or	Skew-	Kurt-
Dataset	N		p		p		p	test	Test	Estimate? *	ness	osis
DOC_m	237	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	1.269	1.922
avslope	859	No	0.0011	No	< 0.0001	No	< 0.0001	Yes	Yes	G	0.4782	0.3732
perRAT	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.100	7.939
perOBM	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.926	13.19
ratOBM	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.229	8.868
RATlintrend	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.236	9.402
Drainden	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	0.5017	0.3765
streamlength	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	11.65	174.6
Catarea	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	12.64	197.6
Lakearea	859	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	G	16.42	288.4
Q	383	No	0.0037	No	0.0339	No	0.0125	Yes	Yes	G	0.1744	-0.4408
DOC	383	No	0.0002	No	< 0.0001	No	< 0.0001	Yes	Yes	G	0.7618	0.6164
avslope	20	Yes	> 0.10	Yes	0.2926	Yes	0.4560	Yes	Yes	G	0.1962	-1.119
catarea	20	No	0.0013	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.997	10.54
catperi	20	Yes	> 0.10	No	0.0020	No	0.0078	Yes	Yes	G	1.562	2.921
areaperi	20	Yes	> 0.10	No	0.0004	No	0.0040	Yes	Yes	G	1.716	4.196
obm	20	No	< 0.0001	No	< 0.0001	No	< 0.0001	Yes	Yes	Exact	2.325	5.435
rat	20	No	0.0004	No	< 0.0001	No	< 0.0001	Yes	Yes	Exact	2.540	6.696
for	20	No	0.0017	No	< 0.0001	No	< 0.0001	Yes	Yes	G	2.901	9.746
forobm	20	No	< 0.0001	No	< 0.0001	No	< 0.0001	No	No	Exact	3.745	14.49
forrat	20	No	0.0063	No	< 0.0001	No	< 0.0001	Yes	Yes	Exact	2.475	7.376
obmav	20	No	< 0.0001	No	0.0050	No	< 0.0001	Yes	Yes	Exact	1.657	1.489
obmmax	20	No	< 0.0001	No	0.0196	No	< 0.0001	Yes	Yes	Exact	1.463	0.7362
obmmin	20	No	< 0.0001	No	0.0042	No	< 0.0001	Yes	Yes	Exact	1.706	1.475
ratav	20	Yes	> 0.10	Yes	0.6922	Yes	0.1641	Yes	Yes	Exact	0.1581	-0.7711
ratmax	20	Yes	> 0.10	Yes	0.4267	Yes	0.1741	Yes	Yes	Exact	0.2464	-0.9728
ratmin	20	No	0.0001	No	0.0023	No	0.0007	Yes	Yes	G	1.644	2.415
spond	20	No	< 0.0001	No	< 0.0001	No	< 0.0001	No	No	Exact	4.301	18.81
wpond	20	No	< 0.0001	No	< 0.0001	No	< 0.0001	No	Yes	Exact	3.884	15.90
rat2	20	Yes	> 0.10	No	0.0170	No	0.0222	Yes	Yes	Exact	1.179	2.135
rat3	20	Yes	> 0.10	No	0.0338	Yes	0.0698	Yes	Yes	G	1.047	1.871
obm2	20	No	< 0.0001	No	0.0053	No	< 0.0001	Yes	Yes	Exact	1.668	1.383
obm3	20	No	0.0009	No	0.0311	No	0.0002	Yes	Yes	Exact	0.6679	-1.408
lakearea	20	Yes	> 0.10	Yes	0.7142	No	0.0337	Yes	Yes	G	-0.1329	-0.7406
lake/cat	20	No	0.0068	No	0.0076	No	0.0074	Yes	Yes	G	1.423	2.039

road	20	Yes	0.0599	Yes	0.0832	No	0.0135	Yes	Yes	G	1.090	0.6016
Strslope	20	Yes	> 0.10	Yes	0,9732	Yes	0,5082	Yes	Yes	G	-0,04942	-0,0215
Strslope_len	20	No	0,0012	No	0,0077	No	0,0005	Yes	Yes	G	1,510	1,660
drainden	20	Yes	0.0659	Yes	0.1049	No	0.0330	Yes	Yes	G	0.9675	0.8405

* Estimate G is Gaussian Approximation,

Sources of data is the program Prism, form GraphPad. Values of zero were ignored by the program during the analysis.

Table App- 6. Measured DOC$_m$ and estimated DOC$_{est}$ from the three models, with the two different RAT layers. Data is first just DOC then log$_{10}$(DOC+1).

				Percentile					Standard		KS	D´A P	SW		Rank	G* or	Skew-	Kurt-
Model	DOC	N	Min	25 %	75 %	Max	Mean	Median	Dev**	Error	***	****	*****	t-test	test	Exact	ness	osis
All lakes with measurements																		
	m	237	1.36	3.1	4	5.31	13.05	4.53	2.003	0.13	No	No	No	Yes	Yes	G	1.263	1.89
m1	est	237	1.075	4.344	5.852	7.646	13.73	6.044	2.393	0.156	Yes	No	No	Yes	Yes	G	0.471	0.1
m3obm	Est	237	0.913	4.13	5.672	7.378	13.57	5.864	2.432	0.158	Yes	No	No	Yes	Yes	G	0.573	0.305
m3lin	Est	237	0.913	4.14	5.679	7.459	13.57	5.899	2.438	0.158	Yes	No	No	Yes	Yes	G	0.552	0.254
m8obm	Est	237	1.099	4.154	5.471	7.086	13.52	5.779	2.299	0.149	Yes	No	No	Yes	Yes	G	0.642	0.486
m8lin	Est	237	1.099	4.154	5.595	7.213	13.52	5.817	2.305	0.15	Yes	No	No	Yes	Yes	G	0.617	0.428
	Log m	237	0.373	0.613	0.699	0.8	1.148	0.718	0.146	0.009	No	Yes	No	Yes	Yes	G	0.337	-0.118
m1	Log est	237	0.317	0.728	0.836	0.937	1.168	0.821	0.157	0.01	No	No	No	Yes	Yes	G	-0.514	0.313
m3obm	Log est	237	0.282	0.71	0.824	0.923	1.163	0.808	0.162	0.011	No	No	No	Yes	Yes	G	-0.483	0.331
m3lin	Log est	237	0.282	0.711	0.825	0.927	1.163	0.81	0.162	0.011	No	No	No	Yes	Yes	G	-0.496	0.339
m8obm	Log est	237	0.322	0.712	0.811	0.908	1.162	0.806	0.153	0.01	Yes	No	No	Yes	Yes	G	-0.397	0.32
m8lin	Log est	237	0.322	0.712	0.819	0.915	1.162	0.808	0.153	0.01	Yes	No	No	Yes	Yes	G	-0.413	0.324
Only RAT wetland on catchments																		
	m	175	1.475	3	4.2	5.55	13.05	4.594	2.115	0.16	No	No	No	Yes	Yes	G	1.26	1.791
m3obm	Est	175	0.913	4.066	5.604	7.476	13.57	5.925	2.589	0.196	Yes	No	No	Yes	Yes	G	0.61	0.175
m3lin	Est	175	0.913	4.066	5.604	7.476	13.57	5.926	2.589	0.196	Yes	No	No	Yes	Yes	G	0.61	0.175
m8obm	Est	175	1.099	4.143	5.509	7.233	13.52	5.852	2.455	0.186	Yes	No	No	Yes	Yes	G	0.669	0.307
m8lin	Est	175	1.099	4.143	5.509	7.233	13.52	5.853	2.455	0.186	Yes	No	No	Yes	Yes	G	0.669	0.307
	Log m	175	0.394	0.602	0.716	0.816	1.148	0.72	0.152	0.012	No	Yes	Yes	Yes	Yes	G	0.334	-0.208
m3obm	Log est	175	0.282	0.705	0.82	0.928	1.163	0.809	0.171	0.013	Yes	No	No	Yes	Yes	G	-0.455	0.287
m3lin	Log est	175	0.282	0.705	0.82	0.928	1.163	0.809	0.171	0.013	Yes	No	No	Yes	Yes	G	-0.455	0.287
m8obm	Log est	175	0.322	0.711	0.814	0.916	1.162	0.808	0.161	0.012	Yes	Yes	Yes	Yes	Yes	G	-0.364	0.255
m8lin	Log est	175	0.322	0.711	0.814	0.916	1.162	0.808	0.161	0.012	Yes	Yes	Yes	Yes	Yes	G	-0.365	0.255
Headwater lakes only																		
	m	117	1.36	3	3.9	5.3	13.05	4.513	2.214	0.205	No	No	No	Yes	Yes	G	1.471	2.232
m1	est	117	1.075	3.643	5.291	7.053	13.73	5.575	2.529	0.234	Yes	No	No	Yes	Yes	G	0.71	0.274
m3obm	Est	117	0.913	3.422	5.146	6.889	13.57	5.377	2.55	0.236	Yes	No	No	Yes	Yes	G	0.794	0.466
m3lin	Est	117	0.913	3.438	5.161	6.921	13.57	5.406	2.554	0.236	Yes	No	No	Yes	Yes	G	0.781	0.43

m8obm	Est	117	1.099	3.616	5.017	6.72	12.7	5.375	2.407	0.223	No	No	No	Yes	Yes	G	0.783	0.463
m8lin	Est	117	1.099	3.616	5.12	6.833	12.7	5.406	2.413	0.223	No	No	No	Yes	Yes	G	0.769	0.423
	Log m	117	0.373	0.602	0.69	0.799	1.148	0.712	0.156	0.014	No	Yes	No	Yes	Yes	G	0.533	0.06
m1	Log est	117	0.317	0.667	0.799	0.906	1.168	0.786	0.171	0.016	Yes	Yes	Yes	Yes	Yes	G	-0.262	-0.112
m3obm	Log est	117	0.282	0.646	0.789	0.897	1.163	0.77	0.177	0.016	Yes	Yes	Yes	Yes	Yes	G	-0.233	-0.109
m3lin	Log est	117	0.282	0.647	0.79	0.899	1.163	0.772	0.177	0.016	Yes	Yes	Yes	Yes	Yes	G	-0.241	-0.101
m8obm	Log est	117	0.322	0.664	0.779	0.888	1.137	0.774	0.166	0.015	Yes	Yes	Yes	Yes	Yes	G	-0.206	-0.08
m8lin	Log est	117	0.322	0.664	0.787	0.894	1.137	0.776	0.166	0.015	Yes	Yes	Yes	Yes	Yes	G	-0.213	-0.076
Headwater lakes, with RAT wetlands on catchments																		
	m	90	1.475	3	4.05	5.305	13.05	4.653	2.317	0.244	No	No	No	Yes	Yes	G	1.43	1.991
m3obm	Est	90	0.913	3.439	5.194	6.989	13.57	5.491	2.664	0.281	Yes	No	No	Yes	Yes	G	0.792	0.355
m3lin	Est	90	0.913	3.439	5.194	6.989	13.57	5.491	2.664	0.281	Yes	No	No	Yes	Yes	G	0.792	0.355
m8obm	Est	90	1.099	3.623	5.113	6.877	12.7	5.474	2.504	0.264	No	No	No	Yes	Yes	G	0.796	0.386
m8lin	Est	90	1.099	3.623	5.113	6.877	12.7	5.474	2.504	0.264	No	No	No	Yes	Yes	G	0.796	0.386
	Log m	90	0.394	0.602	0.703	0.8	1.148	0.722	0.159	0.017	Yes	Yes	No	Yes	Yes	G	0.532	-0.059
m3obm	Log est	90	0.282	0.647	0.792	0.903	1.163	0.776	0.182	0.019	Yes	Yes	Yes	Yes	Yes	G	-0.264	0.004
m3lin	Log est	90	0.282	0.647	0.792	0.903	1.163	0.776	0.182	0.019	Yes	Yes	Yes	Yes	Yes	G	-0.264	0.004
m8obm	Log est	90	0.322	0.665	0.786	0.896	1.137	0.779	0.17	0.018	Yes	Yes	Yes	Yes	Yes	G	-0.219	0.035
m8lin	Log est	90	0.322	0.665	0.786	0.896	1.137	0.779	0.17	0.018	Yes	Yes	Yes	Yes	Yes	G	-0.219	0.035

* G is short for Gaussian Approximation, comes from the Wilcoxon Signed Rank Test.

** Dev = Deviation

*** KS = Kolmogorov-Smirnor normality test

**** D´A P = D'Agostino & Pearson omnibus normality test

***** SW = Shapiro-Wilk normality test

APPENDIX I - RESULTS FROM UNCERTAINTY ANALYSIS – MULTI-MODEL ANALYSIS IN EXCEL.

Table App- 7. Number of regression out of the 10 000 that had an r^2 above 0.75 for both the calibration and validation, with and without duplicates and the difference there between. Percentage of the regressions above 0.75 that were duplicates is also presented.

	Number of	Number of	Number – without duplicates					Percentage Duplicates			
	10 000	10 000			All 10 000 runs				All 10 000 runs		
	0.75	0.75**	0.75	0.75**	10*	7*	5*	0.75	10*	7*	5*
M1	3	1	1	1	5435	5389	5389	66.67	45.65	46.11	46.11
M2	52	5	11	3	5498	5414	5414	78.85	45.02	45.86	45.86
M3	10	14	5	11	5461	5389	5389	50.00	45.39	46.11	46.11
M4	27	26	20	18	5526	5427	5427	25.93	44.74	45.73	45.73
M5	335	312	134	146	5482	5387	5387	60.00	45.18	46.13	46.13
M6	119	121	74	69	5566	5484	5484	37.82	44.34	45.16	45.16
M7	107	96	42	49	5482	5387	5387	60.75	45.18	46.13	46.13
M8	46	43	28	30	5546	5477	5477	39.13	44.54	45.23	45.23

* Number of decimals to which the data was rounded to as duplicates were removed.

** Second simulation of 10 000 runs

Table App- 8. Number of times catchments were used in calibration for the different models, for runs with an r^2 above 0.75 for both calibration and validation. Total is the sum for all models.

		M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**Total**
Number of runs		1	11	5	20	134	74	42	28	315
1	BC1	0	5	3	14	95	66	28	13	224
2	CB1	0	0	0	0	1	2	7	2	12
3	CB2	1	10	5	20	103	55	39	26	259
4	CN1	0	8	0	0	107	54	20	1	190
5	DE11	1	7	5	19	113	64	36	26	271
6	DE10	1	11	5	20	133	74	42	28	314
7	DE5	1	11	4	19	123	68	26	24	276
8	DE6	1	10	5	18	128	70	36	26	294
9	DE8	1	11	5	20	121	70	40	28	296
10	HP3	1	11	4	20	126	68	38	28	296
11	HP3A	1	11	5	20	127	73	42	28	307
12	HP4	1	11	5	18	103	49	38	25	250
13	HP5	1	11	4	17	122	62	37	25	279
14	HP6	1	11	5	18	113	51	32	17	248
15	HP6A	1	10	5	15	111	65	40	24	271
16	PC1	0	5	1	7	61	33	28	14	149
17	RC1	0	8	0	8	94	44	24	11	189
18	RC2	1	8	5	11	55	25	29	26	160
19	RC3	1	3	4	17	69	53	26	25	198
20	RC4	1	3	5	19	105	64	22	23	242

Table App- 9. Ranges of the parameters for each model, (M1, M3 and M8) for the MRW, for both all 10 000 runs and the runs that gave $r^2 > 0.75$.

Model	Parameter	Parameter name	Range [min;max]	Range for $r^2 > 0.75$
M1	Intercept		[6608;10708]	*
	X1	Avslope	[-(980.4;346.7)]	*
M2	Intercept		[6683;10886]	[9885;10560]
	X1	Avslope	[-(980.5;311.2)]	[-(740.1;654:1)]
	X2	Wpond	[-381.9;387:27]	[-381.9;155.3]
M3	Intercept		[3378;10133]	[9104:9932]
	X1	Avslope	[-(866.4;80.79)]	[-(705.6;602.0)]
	X2	perRAT	[-2.080;287.0]	[9.065:49.41]
M4	Intercept		[5232;22682]	[6679;14660]
	X1	Avslope	[-(977.7;331.7)]	[-(803.8;616.3)]
	X2	perFOR	[-146.5;38.88]	[-59.01;35.62]
M5	Intercept		[5909;12332]	[6514;10780]
	X1	Avslope	[-(1058;299.1)]	[-(965.2;370.0)]
	X2	Wpond	[-467.6; 460.4]	[-386.5;383.6]
	X3	PerFORRAT	[-22.34;35.15]	[-3.350;35.15]
M6	Intercept		[2841;11552]	[3763;10520]
	X1	Avslope	[-(939.3;59.68)]	[-(894.2;158.7)]
	X2	PerRAT	[-65.46; 270.5]	[-4.550;180.6]
	X3	PerFORRAT	[-42.54;35.01]	[-25.79;35.01]
M7	Intercept		[3857;11161]	[5058,11160]
	X1	Avslope	[-(925.4;122.5)]	[-(922.3;284.6)]
	X2	PerRAT	[-24.45;263.5]	[-24.45;232.3]
	X3	Wpond	[-392.6;450.7]	[-392.6;367.7]
M8	Intercept		[2144;10395]	[8630:10390]
	X1	Avslope	[-(832.7;72.46)]	[-(739.4;528.0)]
	X2	PerRAT	[-18.40;353.92]	[-16.17;142.1]
	X3	Drainden	[-1.3E6;6.2E5]	[-6.9;2.8]E5

* Only one run gave an r^2 above 0.75 for both datasets.

Table App- 10. Models obtained from all runs when the duplicates have been removed, the statistics r^2 and p. Mean and median parameter values are shown for comparison.

		Calibration		Validation	
	Model equation mean and [median]	r^2	p	r^2	p
M1	y = 9031[9087] – 626.4 [-628.1] · x1	0.517	0.004	0.527	0.223
M2	y = 9165[9220] – 616.3 [-614.7] · x1 - 1058.5 [-105.9] · x2	0.553	0.024	0.686	0.263
M3	y = 7536[7593] – 498.2 [-501.5] · x1 + 68.35 [63.06] · x2	0.573	0.018	0.689	0.270
M4	y = 8886[8591] – 622.8 [-625.0] · x1 + 1.52 [5.73] · x2	0.540	0.027	0.576	0.259
M5	y = 8564[8437] – 575.9 [-561.8] · x1 – 100.5 [-99.29] · x2 + 6.85 [6.63] · x3	0.574	0.058	0.835	0.323
M6	y = 7079[7184] – 469.0 [-469.8] · x1 + 67.68 [60.29] · x2 + 6.27 [6.03] · x3	0.595	0.046	0.838	0.337
M7	y = 7737[7760] – 494.8 [-492.5] · x1 + 66.26 [60.79] · x2 – 106.6 [-110.31] · x3	0.605	0.041	0.842	0.324
M8	y = 7720[7787] – 474.8 [-477.3] · x1 + 62.77 [57.61] · x2 – 180 000 [-1.2E+5] · x3	0.591	0.046	0.836	0.344

Table App- 11. Parameter values for only unique runs with both calibration and validation r^2 above 0.75. The statistical values of r^2 and p are mean values.

		Statistics*				Calibration		Validation	
	Models	**F**	ss_{reg}	ss_{resid}	se_y	r^2	p	r^2	p
M1	y = 10080 – 669.9 · x1	47.81	5.3E7	1.4E7	1054	0.79	0.0000	0.83	0.033
M2	y = 10210[10270*] – 716.9[-730.5] · x1 – 230.9[118.4] · x2	40.95	5.5E7	1.8E7	1160	0.77	0.0008	0.84	0.143
M3	y = 9712[9477] – 670.0[-661.5] · x1 + 22.15[30.81] · x2	19.31	5.7E7	1.8E7	1215	0.76	0.0009	0.96	0.034
M4	y = 9766[9467] – 702.6[-691.7] · x1 + 3.40[5.601] · x2	19.68	6.0E7	1.8E7	1236	0.77	0.0008	0.80	0.311
M5	y = 7923[7930] – 527.4[-528.7] · x1 – 160.1[151.7] · x2 + 21.10[21.23] · x3	13.45	6.5E7	1.8E7	1277	0.78	0.0043	0.92	0.274
M6	y = 6589[7008] – 446.8[-479.1] · x1 + 47.41[40.68] · x2 + 22.18[20.23] · x3	12.58	6.5E7	1.9E7	1312	0.78	0.0046	0.92	0.258
M7	y = 9067[8897] – 648.5[-627.8] · x1 + 54.34[61.89] · x2 – 79.70[-154.8] · x3	12.68	6.3E7	1.8E7	1290	0.77	0.0051	0.92	0.268
M8	y = 9711[9620] – 611.5[-612.7] · x1 + 23.32[29.84] · x2 – 311100[-334000] · x3	12.32	6.1E7	1.8E7	1286	0.77	0.0052	0.96	0.188

* The values in [] were obtained during a second 10 000 runs simulation, made to gain the values of F, ss_{reg}, ss_{resid}, and se_y for the mean models. It was also made to see if the values would differ dramatically in a second term of runs. For model M1 they values stayed the same, for other models the difference could be quite high.

APPENDIX J - RESULTS FROM SENSITIVITY AND MULTI-MODEL ANALYSIS IN EXCEL.

Table App- 12. Sensitivity analysis made with regression coefficients and intercept from a regression made with all 20. The mean values and range for DOC estimated for each model with one parameter at the time multiplied by a random number between 0.75-1.25 (±25 %). 10 000 runs were made.

		X1				X2				X3			
	Ref	Mean	Min	Max	%	Mean	Min	Max	%	Mean	Min	Max	%
M1	4985.11	4987.57	3880.92	6089.18	22.149								
M2	4591.11	4593.77	3397.86	5784.22	25.989	4589.16	4434.40	4747.80	3.413				
M3	4970.22	4972.15	4108.53	5831.82	17.336	4972.16	4813.91	5126.55	3.145				
M4	5163.68	5166.08	4090.38	6236.87	20.785	5211.87	4972.59	5589.72	5.843				
M5	4666.56	4668.95	3598.71	5734.29	22.882	4664.91	4533.15	4799.95	2.859	4654.95	4562.52	4727.14	1.772
M6	5000.51	5002.02	4324.09	5676.85	13.526	5002.36	4851.55	5149.48	2.979	4980.14	4818.04	5106.75	2.909
M7	4602.84	4605.02	3625.77	5579.80	21.227	4604.52	4467.08	4738.61	2.950	4604.01	4455.94	4749.68	3.191
M8	4944.31	4946.30	4055.19	5833.33	17.982	4946.40	4775.89	5112.75	3.406	4943.97	4901.99	4986.64	0.856

Table App- 13. Sensitivity analysis made with regression coefficients and intercept from the mean regression of the runs above $r^2 = 0.75$. The mean values and range for DOC estimated for each model with one parameter at the time multiplied by a random number between 0.75-1.25 (±25 %). 10 000 runs were made.

		X1				X2				X3			
	Ref	Mean	Min	Max	%	Mean	Min	Max	%	Mean	Min	Max	%
M1	5740.74	5743.87	4657.04	6824.27	18.876								
M2	5145.72	5024.24	3891.12	6150.65	22.501	4877.66	5164.32	5164.30	2.855				
M3	5588.89	5639.33	4558.84	6713.42	19.114	5597.39	5667.49	5675.01	0.622				
M4	5502.98	5549.33	4403.05	6688.84	20.608	5542.69	5555.57	5556.06	0.116				
M5	5229.12	5176.54	4279.60	6068.17	17.285	5091.10	5241.53	5256.82	1.456	5151.26	4960.36	5286.17	3.180
M6	5216.52	5227.92	4506.71	5944.86	13.760	5124.63	5327.02	5327.01	1.936	5196.47	4949.39	5371.07	4.086
M7	3414.90	5297.39	4245.29	6343.26	19.814	5041.43	5205.70	5294.36	1.603	5294.44	5037.60	5216.19	1.742
M8	5473.29	5476.15	4484.07	6462.37	18.073	5437.59	5528.54	5529.59	0.829	5473.57	5361.85	5604.25	2.210

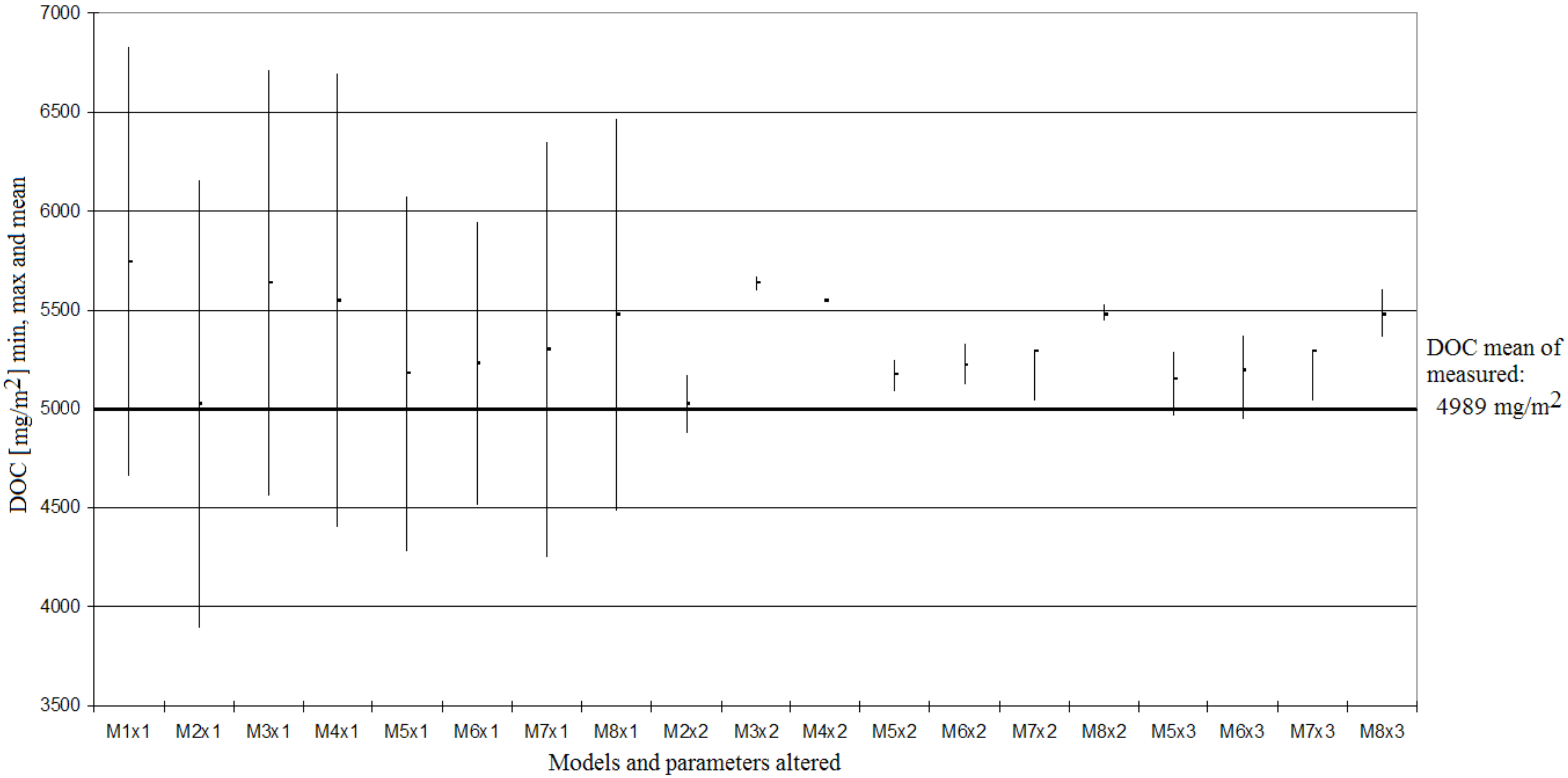
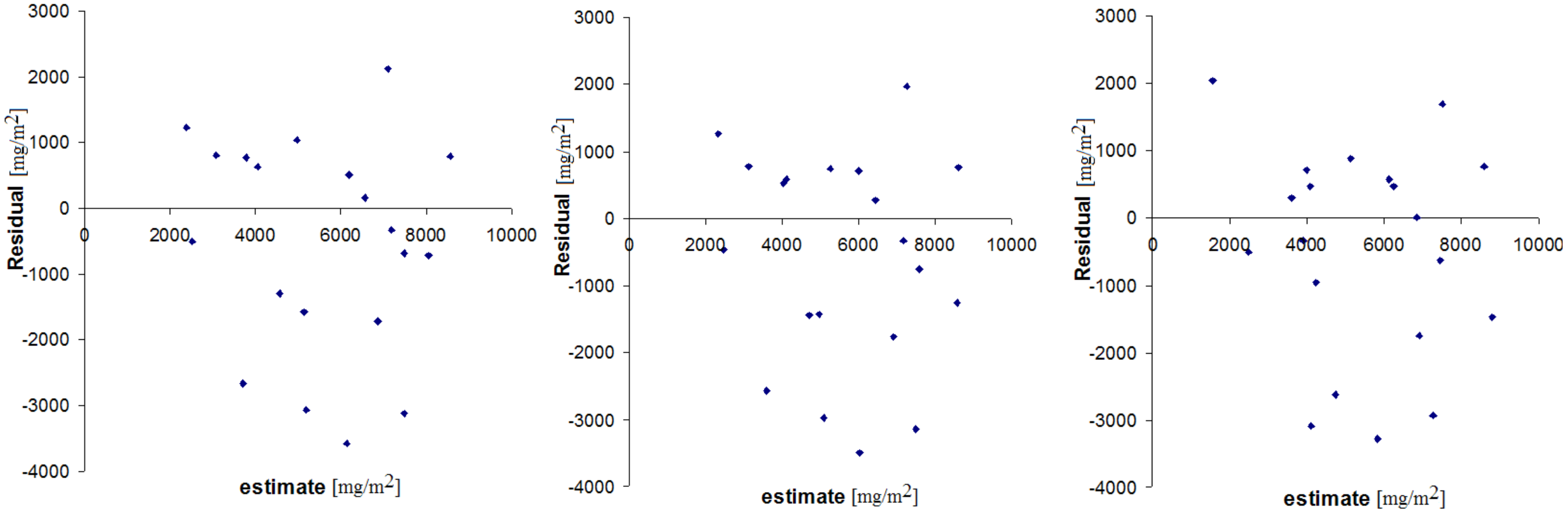


Figure App- 10. Ranges from sensitivity analysis for all models, with the different parameters, x1, x2 and x3 changed one at the time. Max and min are at ends of the lines and mean is marked. The thicker line is the mean of the measured DOC for all subcatchments. The models have the highest range for average slope, but for some other values the range does not even cover the mean value.

APPENDIX K – RESIDUALS ANALYSIS OF THE THREE CHOSEN MODELS FOR DORSET



a) Model M1. Residual plot x against $\hat{Y}$.

c) Model M3. Residual plot x against $\hat{Y}$.

c) Model M8. Residual plot x against $\hat{Y}$.

Figure App- 11. The models residuals plotted against the parameter values and the interactions between the parameters (only for models M3 and M8).

APPENDIX L –OPTIMIZATION OF V_U AND V_L IN THE LAKE DOC MODEL ON THE MUSKOKA RIVER WATERSHED

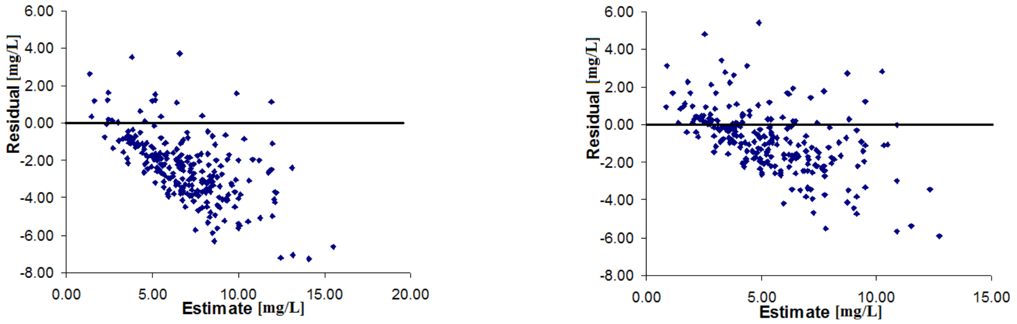
Table App- 14. The two v´s for the Lake model were changed simultaneously or one at the time – from the starting value of 3 ± 1 [m/yr]. The absolute average deviations, for the three models, with the linear trend to fill up the gap in Ducks unlimited wetlands and only those that had wetlands.

						v_u	4	2	3	3
Models	N	v	3	4	2	v_l	3	3	4	2
M1	237		3.66	2.94	4.65		1.79	2	1.53	2.47
M3	237		1.78	1.39	2.47		1.68	1.89	1.46	2.33
M3RAT	175		1.84	1.45	2.54		1.75	1.96	1.52	2.4
M8	237		1.67	1.29	2.37		1.58	1.78	1.35	2.23
M8RAT	175		1.74	1.36	2.44		1.65	1.85	1.42	2.3

Table App- 15. Results, regression coefficients, r^2, r and other statistics from comparison to measured data and the estimated. Based on v_u and $v_l = 3$. Results are from S-PLUS and Excels LINEST. The linear trend between the wetland types was used.

Model	Equation	r^2	r	t0	t1
M1	$DOC_m = 0.836 + 2.257\ DOC_{est}$	0.49	0.7	0.622	0.047
M3	$DOC_m = 0.867 + 1.971\ DOC_{est}$	0.508	0.713	0.544	0.048
M3_RAT	$DOC_m = 0.870 + 1.928\ DOC_{est}$	0.505	0.711	0.638	0.057
M8	$DOC_m = 0.833 + 2.041\ DOC_{est}$	0.525	0.724	0.523	0.043
M8_RAT	$DOC_m = 0.834 + 2.023\ DOC_{est}$	0.516	0.718	0.628	0.051
Old*	$DOC_m = 0.979 + 0.357\ DOC_{est}$	0.5	0.707	0.17	0.063
Old_RAT	$DOC_m = 1.021 + 0.123\ DOC_{est}$	0.517	0.719	0.047	0.077

* Old means the old mass balance model from Dillon and Molot (1997b) was used.



a) v_u and $v_l = 2$ c) v_u and $v_l = 4$

Figure App- 12. Model M1 all 237 catchments with measurements.

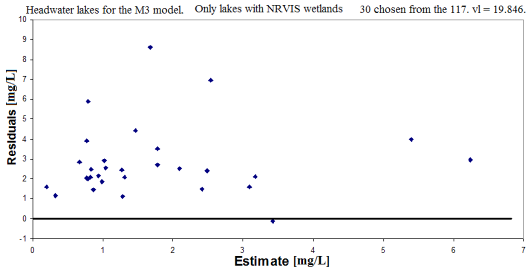


Figure App- 13. Plot of 30 residuals towards DOC_{est} values for one set of values of v_u and v_l obtained from a minimum of absolute deviations of 1000 runs in Excel.

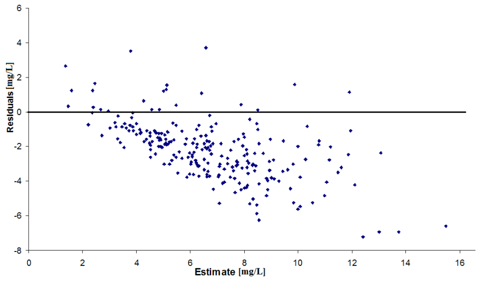


Figure App- 14. Model M1, residual plot using loss coefficients at maximum r^2, from runs in Excel.

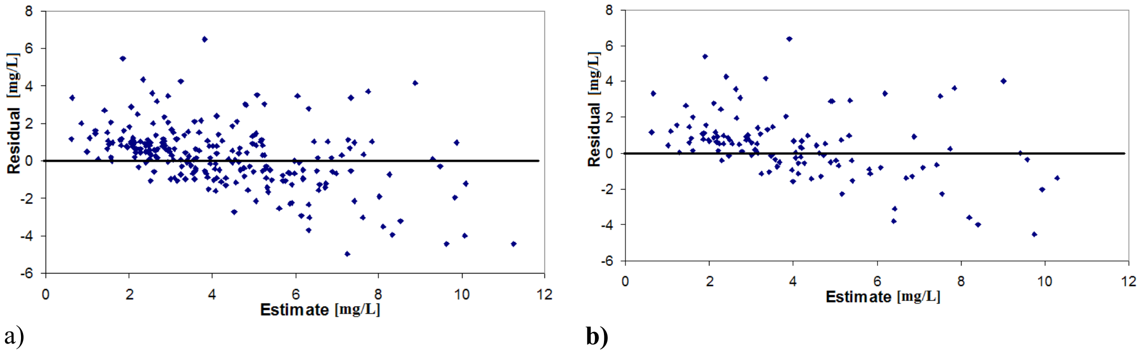


Figure App- 15. Residuals and estimated DOC values for M1, of the optimized values from Crystal ball. a) all 237 measured values, and b) all headwater lakes.

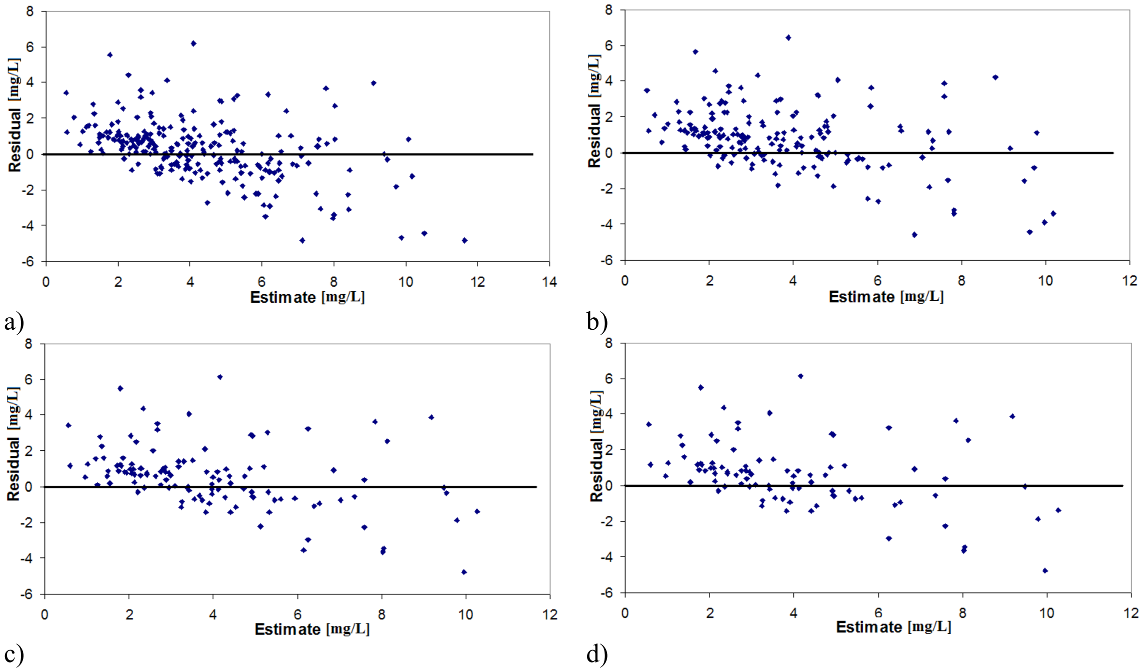
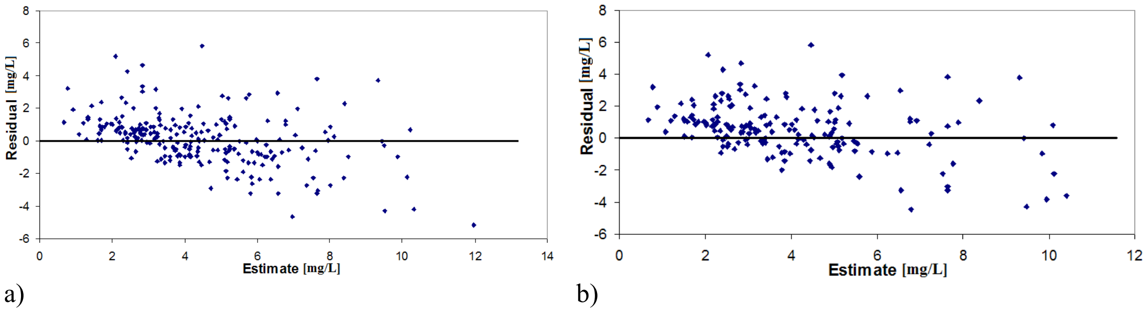
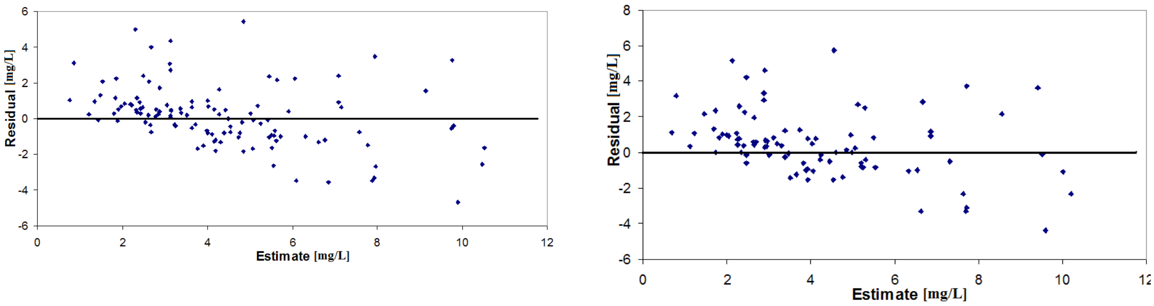


Figure App- 16. Residuals and estimated DOC values for M3, of the optimized values from Crystal ball. a) all 237 measured values, b) all lakes with Ducks unlimited wetlands, c) all headwater lakes and d) all headwater lakes with Ducks unlimited wetlands.





c) **d)**

Figure App- 17. Residuals and estimated DOC values for M8, of the optimized values from Crystal ball. a) all 237 measured values, b) all lakes with Ducks unlimited wetlands, c) all headwater lakes and d) all headwater lakes with Ducks unlimited wetlands.

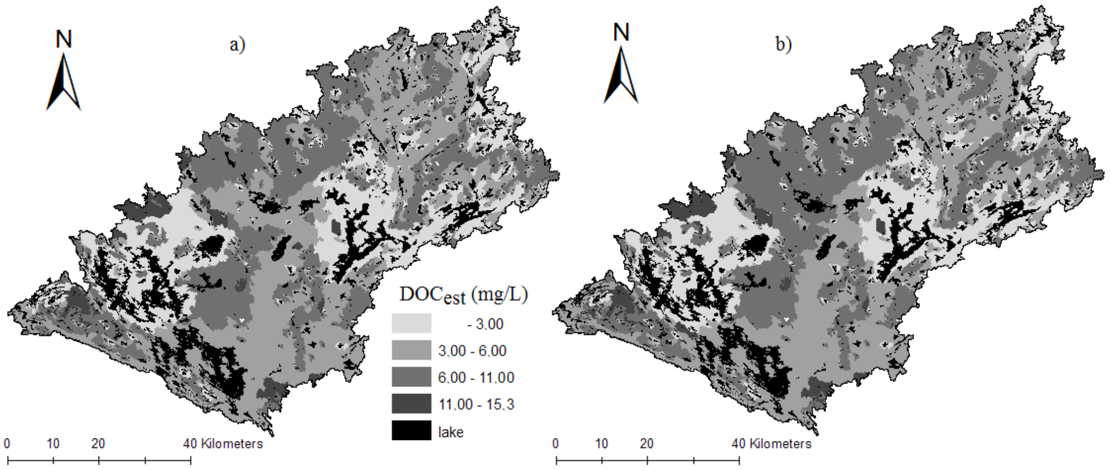


Figure App- 18. Estimated DOC lake concentrations for all 859 lakes, from model a) M1 and b) M3, here plotted on the catchments for better clarity.

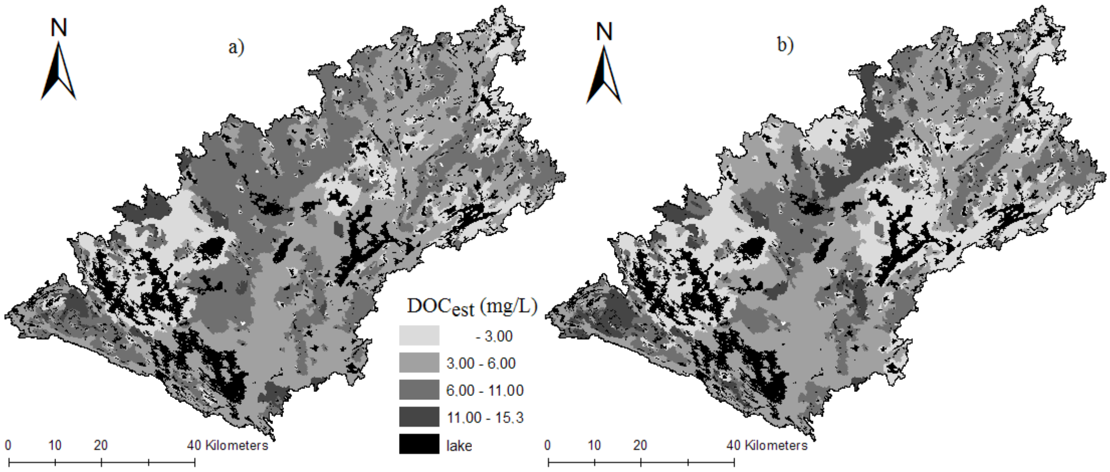


Figure App- 19. Estimated DOC lake concentrations for all 859 lakes, from a) model M8 and b) the old peat model, here plotted on the catchments for better clarity.

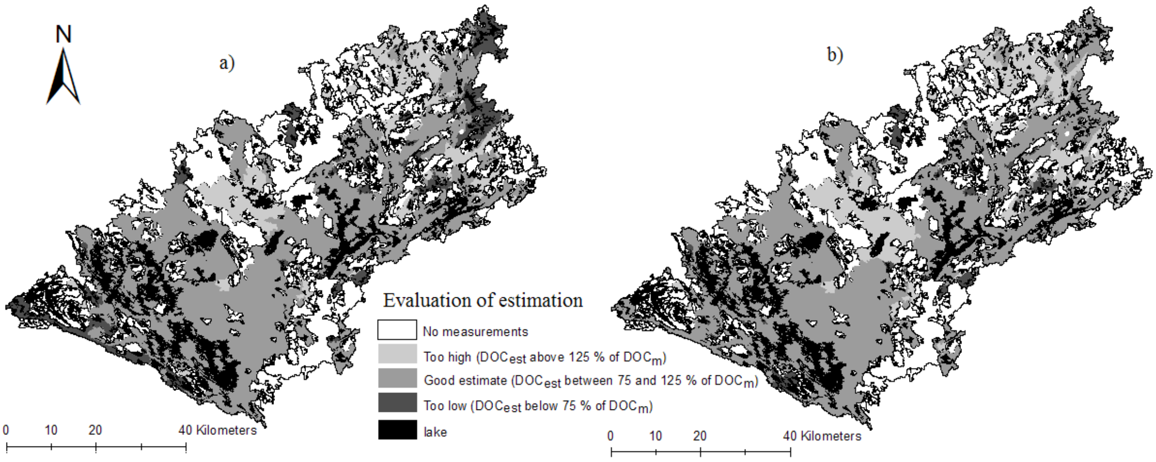

N
a)
b)
Evaluation of estimation
No measurements
Too high (DOC$_{est}$ above 125 % of DOC$_{m}$)
Good estimate (DOC$_{est}$ between 75 and 125 % of DOC$_{m}$)
Too low (DOC$_{est}$ below 75 % of DOC$_{m}$)
lake
0 10 20 40 Kilometers
0 10 20 40 Kilometers

Figure App- 20. Estimates from model a) M1 and b) M8, all 237 with measurements, termed as good, too high or too low as well as the parts where DOC have not been measured, plotted on the catchments for better clarity.

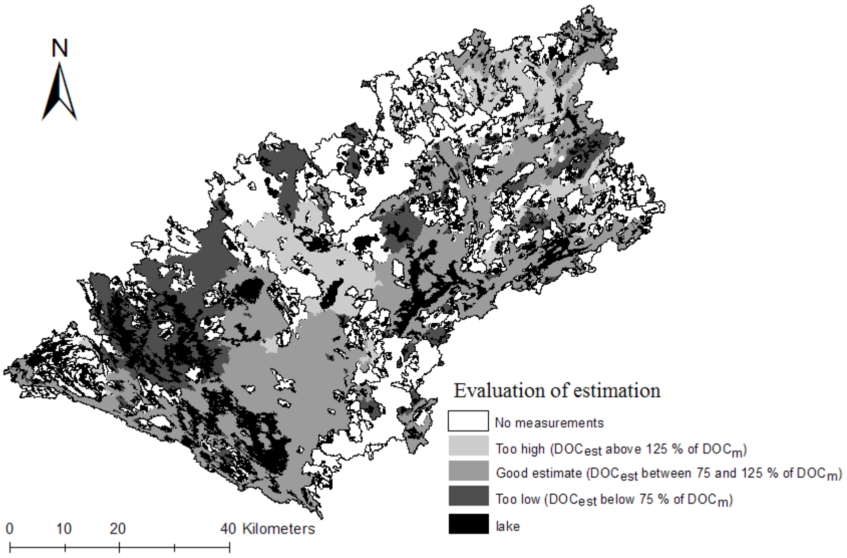

N
Evaluation of estimation
No measurements
Too high (DOC$_{est}$ above 125 % of DOC$_{m}$)
Good estimate (DOC$_{est}$ between 75 and 125 % of DOC$_{m}$)
Too low (DOC$_{est}$ below 75 % of DOC$_{m}$)
lake
0 10 20 40 Kilometers

Figure App- 21. Estimates from the old peat model, all 237 with measurements, termed as good, too high or too low as well as the parts where DOC have not been measured, plotted on the catchments for better clarity.

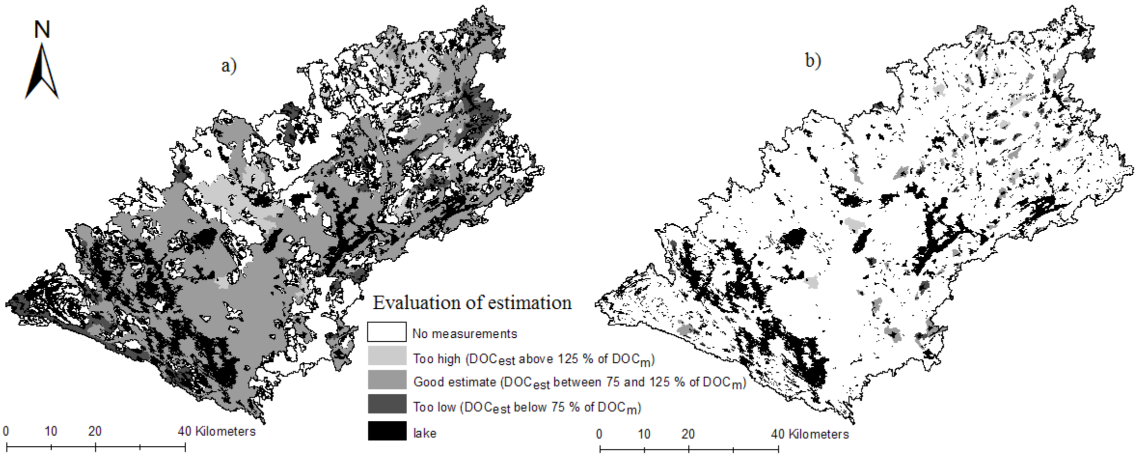

N
a)
b)
Evaluation of estimation
No measurements
Too high (DOC$_{est}$ above 125 % of DOC$_{m}$)
Good estimate (DOC$_{est}$ between 75 and 125 % of DOC$_{m}$)
Too low (DOC$_{est}$ below 75 % of DOC$_{m}$)
lake
0 10 20 40 Kilometers
0 10 20 40 Kilometers

Figure App- 22. Estimates from model M3,a) all 237 with measurements and b) all 117 headwater lakes with measurements, termed as good, too high or too low as well as the parts where DOC have not been measured, plotted on the catchments for better clarity.

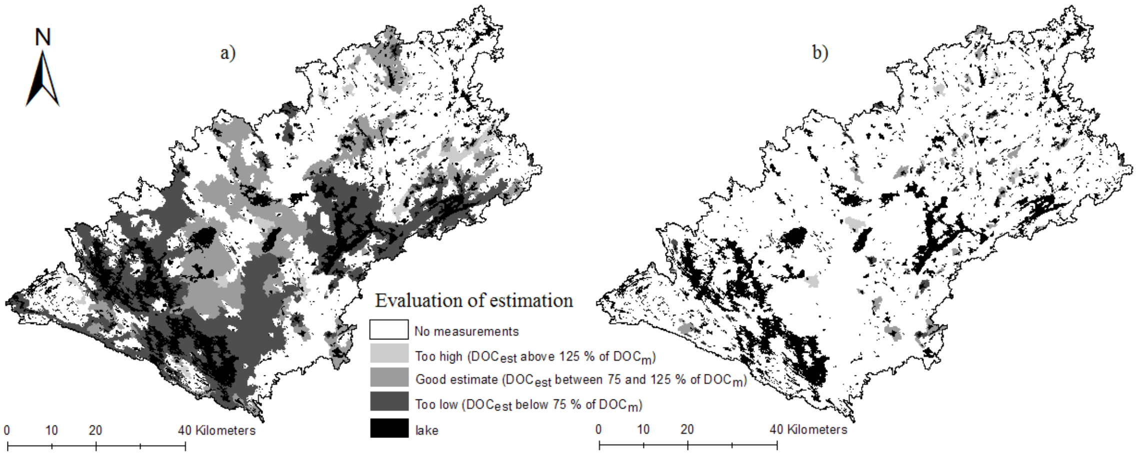

N
a)
b)
Evaluation of estimation
No measurements
Too high (DOC_est above 125 % of DOC_m)
Good estimate (DOC_est between 75 and 125 % of DOC_m)
Too low (DOC_est below 75 % of DOC_m)
lake
0 10 20 40 Kilometers
0 10 20 40 Kilometers

Figure App- 23. Estimates from model M3, a) all 175 lakes with measurements and Ducks unlimited wetlands and b) all associated 90 headwater lakes, termed as good, too high or too low as well as the parts where DOC have not been measured, plotted on the catchments for better clarity.

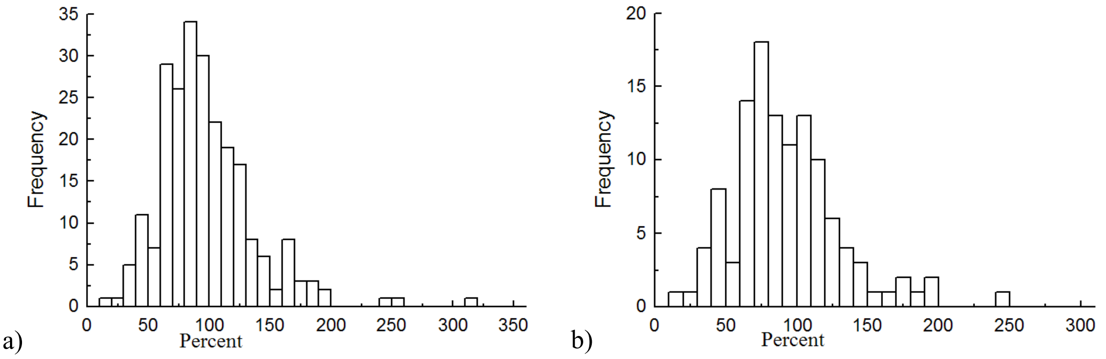


Figure App- 24. Histogram over the distribution of Percentage difference for estimated values from the measured, for model M1, a) all 237 lakes with measurements and b) headwater lakes.

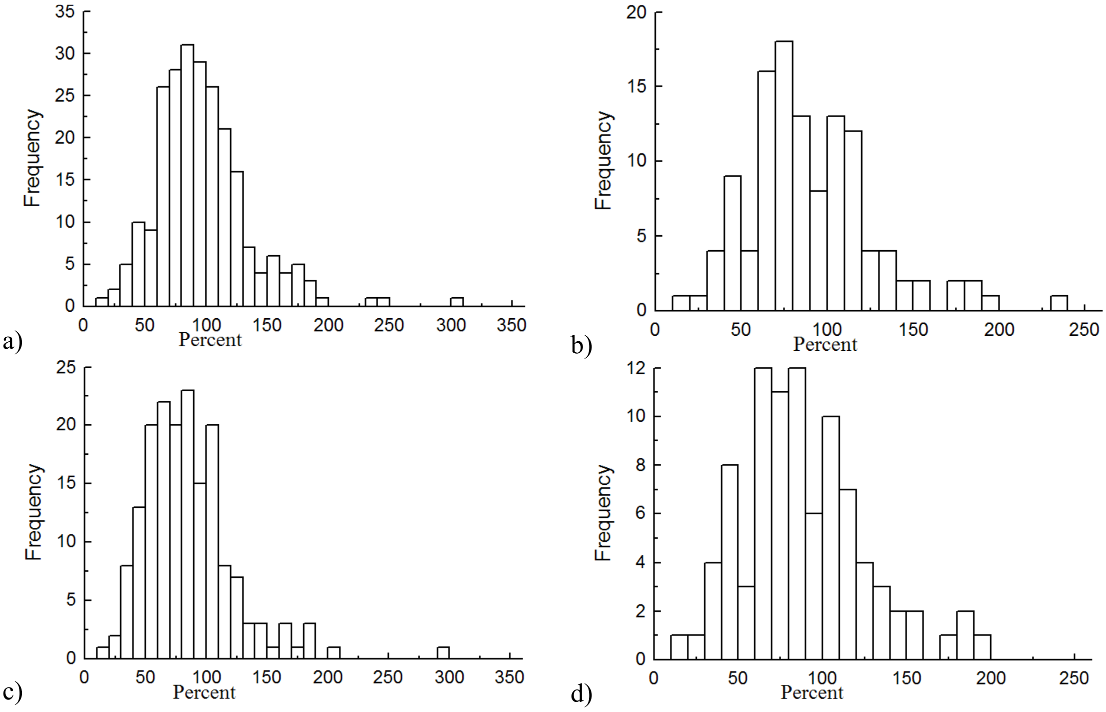


Figure App- 25. Histogram over the distribution of Percentage difference for estimated values from the measured, for model M3, a) all 237 lakes with measurements, b) headwater lakes, c) lake with Ducks unlimited wetlands and d) those in c that are headwater lakes.

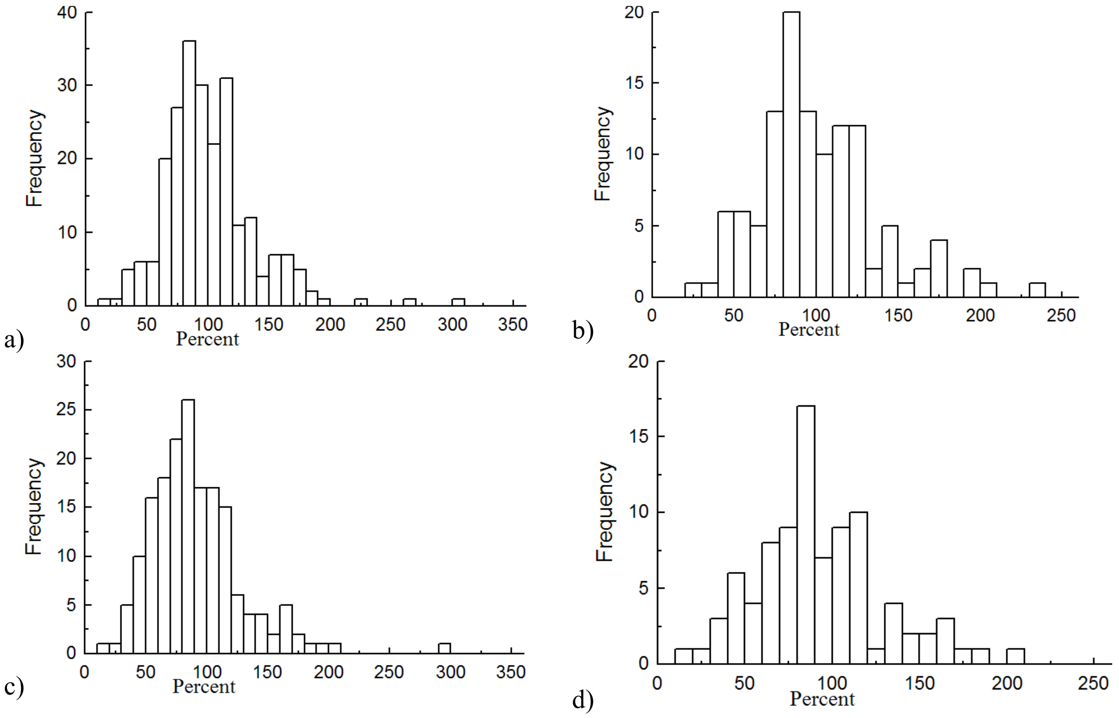


Figure App- 26. Histogram over the distribution of Percentage difference for estimated values from the measured, for model M8, a) all 237 lakes with measurements, b) headwater lakes, c) lake with Ducks unlimited wetlands and d) those in c that are headwater lakes.

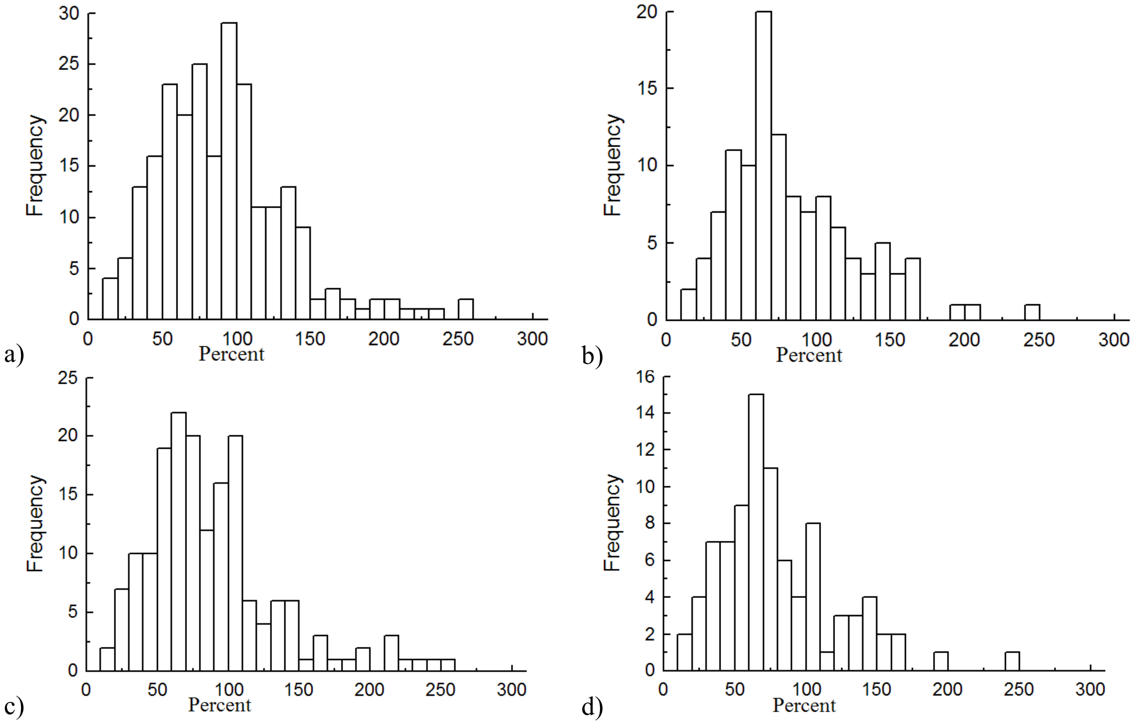


Figure App- 27. Histogram over the distribution of Percentage difference for estimated values from the measured, for the old peat model, a) all 237 lakes with measurements, b) headwater lakes, c) lake with Ducks unlimited wetlands and d) those in c that are headwater lakes.

APPENDIX M - RESULTS FROM UNCERTAINTY- AND SENSITIVITY ANALYSIS – MULTI-MODEL ANALYSIS ON DOC/Q.

Table App- 16. Results from multi-model regression for the same eight models as for DOC, with fifteen subcatchments used for calibration and the remaining five for validation. 10 000 runs were made and the result here is the mean values for all models with r^2 for both calibration and validation > 0.75. Range is also shown as [min;max].

	% Dup-		Calibration		Validation	
	licates*	Models	**r^2**	**p**	**r^2**	**p**
M1	46.13	y = 18130[12857;19721] – 1248[-(1720.1;752.1)] · x1	0.78	0.000	0.79	0.044
M2	45.98	y = 18290[13122;20018] – 1344[-(1734.9;762.3)] · x1 – 339.1[-677.2;718.4] · x2	0.78	0.001	0.88	0.108
M3	45.66	y = 18190[8835;19775] – 1321[-(1646.0;429.0)] · x1 + 16.17[-34.74;398.9] · x2	0.78	0.001	0.89	0.106
M4	45.35	y = 17620[10339;37234] – 1323[-(1636.3;721.4)] · x1 + 6.08[-201.5;70.31] · x2	0.78	0.001	0.80	0.378
M5	46.07	y = 16420[11560;2270510– 1171[-(1746;668.3)] · x1 – 287.2[-757.9;588.6] · x2 + 17.59[-1.28;58.7] · x3	0.79	0.004	0.92	0.266
M6	46.23	y = 14770[7401;22470] – 1058[-(1993;335.5)] · x1 + 36.12[-196.5;364.1] · x2 + 24.63[-66.28;57.32] · x3	0.78	0.004	0.91	0.272
M7	46.3	y = 17480[9270;21320] – 1257[-(1745;433.6)] · x1 + 34.55[-94.36;403.2] · x2 – 322.0[-704.3;888.3] · x3	0.78	0.004	0.92	0.264
M8	46.06	y = 18160[7575;20046] – 1199[-(1574.6;426.9)] · x1 + 13.85[-63.27;476.7] · x2 – 5.8[-(22;5.7)]10^5 · x3	0.78	0.005	0.92	0.265

* seven decimals were used to roundup data before VBA code to remove duplicates was run.

Table App- 17. Mean and range for the DOC estimates, results for the sensitivity analysis ±25 %, for models above.

	X1				X2				X3			
	Mean	Min	Max	%	Mean	Min	Max	%	Mean	Min	Max	%
M1	5743.87	8027.2	12065.7	20.10								
M2	5024.24	6799.2	11146.5	24.22	5021.85	8818.7	9127.6	1.72				
M3	5639.33	7650.5	11926.3	21.84	5635.97	9749.6	9820.1	0.36				
M4	5549.33	7425.6	11705.7	22.37	5547.31	9438.5	9605.3	0.88				
M5	5176.54	7266.0	11054.0	20.68	5174.46	9029.5	9267.1	1.30	5151.26	8950.0	9270.7	1.76
M6	5227.92	7732.0	11155.6	18.13	5225.22	9356.8	9531.3	0.92	5196.47	9149.5	9598.7	2.40
M7	5297.39	7049.2	11117.3	22.39	5294.36	9000.2	9160.5	0.88	5294.44	8937.0	9230.2	1.61
M8	5476.15	7653.6	11533.2	20.22	5472.95	9560.3	9627.2	0.35	5473.57	9358.5	9828.9	2.45