



---

Sveriges  
lantbruksuniversitet

# Prediktion av sjödjup och sjövolym från topografiska data

av

Sebastian Sobek, Jakob Nisell and Jens Fölster



# Prediktion av sjödjup och sjövolym från topografiska data

av

Sebastian Sobek, Jakob Nisell and Jens Fölster

Institutionen för vatten och miljö, SLU

Box 7050

750 07 Uppsala

[www.ma.slu.se](http://www.ma.slu.se)

*Tryck:* Institutionen för vatten och miljö, SLU, årtal (2010)

ISSN 1403-977X

## Sammanfattning

Vi utvecklade en statistisk modell med målet att kunna prediktera sjödjup och sjövolym för alla svenska sjöar enbart utgående från kartbaserade data. Sjövolym kunde förklaras mycket väl utifrån sjöarea och den maximala lutningen i en 50 m bred buffertzona från strandlinjen ( $R^2=0.92$ ). Däremot kunde varken det maximala djupet eller medeldjupet modelleras utifrån kartbaserade data ( $R^2<0.4$ ). Trots den höga förklaringsgraden (92%) är prediktionen av volym för en enskild sjö behäftad med en stor statistisk osäkerhet. Däremot blir det statistiska felet mycket mindre om man predikterar den genomsnittliga volymen för minst 5 sjöar åt gången. Volymmodellen kan därför användas för att prediktera den totala sjövolymen, och i förlängning vattnets uppehållstid för avrinningsområden som innehåller minst 5 sjöar.



## Inledning

En sjös vattendjup och volym bestämmer många av sjöns egenskaper och funktioner. Djup och volym är avgörande för vattnets uppehållstid, och därmed för effekten av sjöprocesser såsom primärproduktion, nedbrytning av organiskt material, näringsomsättning, och sedimentation på vattenkvalitén (Algesten et al. 2004, Hamilton and Lewis 1990, Jeppesen et al. 2005). Vidare är sjöbassängens form viktig för resuspension av sediment (Håkanson 2004), och därmed för sjöns näringsförhållanden och primärproduktion (Hellström 1991). Dessutom påverkar bassängens form skiktningförhållanden, vilka i sin tur har stor inverkan på bottenvattnets syrekoncentration och rörligheten av fosfor i sedimenten (Jeppesen et al. 2005).

Med tanke på hur viktiga sjödjup och sjövolym är för sjöns egenskaper, är det förvånansvärt att dessa grundläggande parameter är okända för de flesta svenska sjöarna. Bara knappt 8000 av de över 100 000 sjöarna i svenska sjöregistret har uppgift om djup eller volym! Detta beror i huvudsak på att det är tidskrävande att loda en sjö, även med tillgång till GPS och ekolod. Därför är det av intresse att hitta statistiska samband, utifrån vilka man skulle kunna prediktera en sjös djup och volym. Håkanson och Peters (1995) visade att sjövolymen kan bestämmas utifrån sjöarealen och den maximala lutningen i tillrinningsområdet. Denna formel är dock inte användbar för alla sjöar, eftersom tillrinningsområdet oftast inte är känt. Det finns visserligen GIS-rutiner som kan bestämma ett tillrinningsområde utifrån en digital höjddatamodell, men dessa rutiner ger inte tillförlitliga resultat i låglänta områden, där höjdskillnaderna i landskapet är mindre än den vertikala upplösningen i tillgängliga höjddatabaser.

Vi försökte därför att hitta statistiska samband mellan djup och volym av sjöar och kartbaserad information om topografi, markanvändning och marktyp. Målet var att prediktera sjödjup och volym för alla sjöar i sjöregistret. Detta mål lyckades vi inte uppnå, eftersom den statistiska osäkerheten i prediktionen för enstaka sjöar var för stor. Däremot kom vi fram till statistiska samband som möjliggör att uppskatta den totala sjövolymen i ett tillrinningsområde, om det innehåller minst 5 sjöar.

## Metoder

Vi utgick ifrån svenska sjöregistret ([www.smhi.se](http://www.smhi.se)) för att undersöka om det finns statistiska samband mellan maximala djupet ( $D_{\max}$ ), medeldjupet ( $D_{\text{med}}$ ) och volymen ( $V$ ) av en sjö och kartbaserade parametrar. Sjöregistret innehåller uppgifter om maximaldjup för 7383 sjöar, och om volym (och därmed medeldjup, som beräknas som volym / area) för 6618 sjöar. För dessa sjöar finns dessutom uppgifter om area ( $A$ ) och höjd över havet (höj) i sjöregistret. Vi uteslöt stora sjöar ( $>10 \text{ km}^2$ ) eftersom djup och volym av dessa sjöar är kända. Felaktiga area- eller volymdata identifierades genom att jämföra beräknat medeldjup med angiven medeldjup, och uteslöts. Totalt ingick 6943 sjöar i analysen. Dessa är belägna i olika naturgeografiska regioner av landet, och representerar därför mycket olika sjötyper (Tabell 1). Det förefaller dock att fjällsjöarna förmodligen är underrepresenterade i datamängden. Sjöarna är mycket olika med avseende på storlek, djup, och volym, men de flesta är relativt små och grunda (medianen i  $A$ :  $0,21 \text{ km}^2$ ;  $V$ :  $0,57 \text{ Mm}^3$ ;  $D_{\max}$ :  $3.4 \text{ m}$ ).

**Tabell 1.** Fördelning av sjöar i olika naturgeografiska områden.

Naturgeografiskt område	N
Fjällområdet	79
Mellersta och norra skogsområden	3596
Norra Norrlands kust och slätt	102
Förfjällens slättområde	38
De stora slättarna	247
Västkusten	282
Höglandets sluttningsområden	1752
Småländska höglandet	751
Öland och Gotland	10
Skåne	51
Kalmarsslätten	12

### Metod för insamling av statistik för sjöns närområde

Geometrin för sjöarna hämtades dels från SMHI:s GIS-skikt med vattenförekomster framtagna utifrån kartskalet 1:250 000 och dels från vägkartan 1:100 000 där data inte fanns i SMHI:s skikt. För varje sjö har fem buffertzoner skapats, 50, 100, 200, 500 och 1000 m från strandlinjen. För varje buffertzonen har statistik om marktyp, jordart, lutning och höjdförhållanden beräknats. Statistiken har lagrats i en textfil tillsammans med information om sjöns identitetskod, geometriska area och omkrets,

De topografiska parametrarna omfattar medelvärdet, maximalvärdet, minimalvärdet och standardavvikelsen av lutningen ( $lut$ ) och höjden över havet ( $höj$ ) i varje av dessa buffertzoner. Dessa topografiska parametrar benämndes enligt parameterstatistik, buffertzon, t ex  $lut_{\max 50}$  för maximala lutningen i en 50 m buffertzon från strandlinjen, eller  $höj_{\text{stav}500}$  för standardavvikelsen av höjden i en 500 m buffertzon från strandlinjen. Marktyper och markanvändning omfattar 11 olika klasser som extraherades genom att generalisera indelningen i Svensk marktäckedata (Engberg, 2002). Även en generalisering med 17 markanvändningstyper gjordes. Sjöns omkrets användes för att beräkna  $s_k$  "shoreline development" eller flikighet, dvs. sjöns avvikelse från cirkulär form, enligt Håkanson (1981).



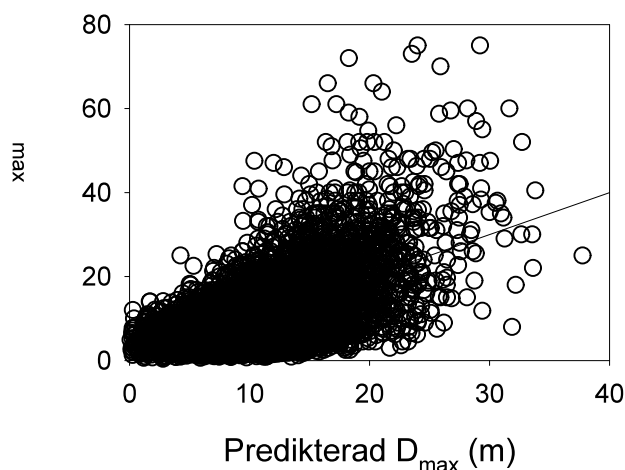
## Statistik

Den statistiska analysen genomfördes sedan i två steg. Först användes multivariata statistiska metoder såsom PCA (principal components analysis) och PLS (partial least squares regression) för att upptäcka den grundläggande korrelationsstrukturen i datamängden. Både PCA och PLS identifierar osynliga ("latenta") gradienter i datamängden, längs med vilka en stor del av variansen i datamängden kan förklaras. Dessa till en början osynliga gradienter kan därmed synliggöras i enkla diagram, som visar huruvida olika variabler är positivt, negativt, eller inte alls relaterade till varandra. Skillnaden mellan PCA och PLS är att en PCA visar korrelationer mellan X-variabler, medan en PLS, som är en regressionsmetod, visar hur olika X-variabler bidrar till att förklara variansen i en Y-variabel. Alla PCA- och PLS-analyser genomfördes i Simca-P+ 12.0 (Umetrics, Umeå).

I ett nästa steg användes sedan de variabler, som i PCA- och PLS-analyserna visade sig vara relaterade till sjödjup och sjövolym, som prediktorer i multipla linjära regressioner. Enbart sinsemellan oberoende variabler användes som prediktorer i multipla linjära regressioner, för att tillfredställa detta antagande i metoden. För dessa analyser användes JMP 8.0. Alla variabler som var tydligt snedfördelade (skewness >2, min/max <0.1) transformerades med naturlig logaritm för att erhålla normalfördelning.

## Resultat och Diskussion

Bara en begränsat andel av variabiliteten i sjödjup kunde förklaras utifrån kartbaserade parametrar. En PLS-regression förklarade 43% av variansen i  $D_{\max}$ , med sjöarea, omkrets, maximala lutningen i sjöns omgivning och strandlinjens flikighet som viktigaste prediktorer (resultat visas ej). De viktigaste och sinsemellan oberoende av dessa variablerna, A och  $\text{lut}_{\max 50}$ , lyckades dock bara förklara 36% procent av variansen i  $D_{\max}$  (Figur 1). Vi drar därifrån slutsatsen att en sjös maximala djup inte kan predikteras från enbart kartbaserade data.



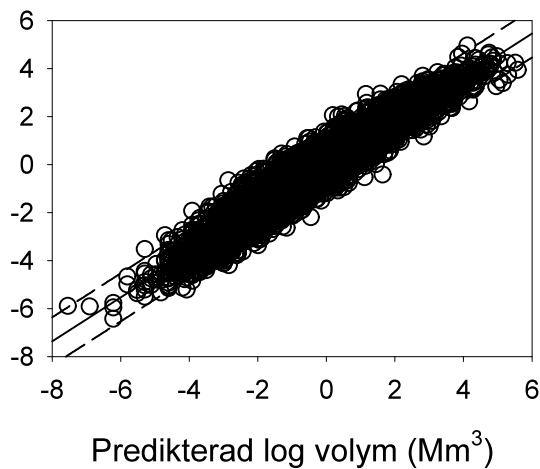
**Figur 1.** Maximaldjup ( $D_{\max}$ ) predikerad från sjöarea och maximala lutningen i en 50 m buffertzona från strandlinjen, mot observerat maximaldjup. Linjen visar följande regression:  $D_{\max} = 7,09 + 1,69 * \log A + 0,660 * \text{lut}_{\max 50}$ ;  $R^2=0,36$ ;  $n=6772$ ;  $p<0,0001$ .

Liknande resultat erhöles för medeldjupet ( $D_{med}$ ).  $D_{med}$  uppvisade visserligen en stark korrelation med  $D_{max}$  ( $R^2=0,78$ ; resultat visas ej), men inte med kartbaserade parametrar ( $D_{med} = 2,2 + 0,33 * \log A + 0,24 * \text{lut}_{max50}$ ;  $R^2=0,31$ ; resultat visas ej). Inte heller en sjös medeldjup kan därför predikteras utifrån kartbaserade parametrar.

Sjövolymen, å andra sidan, kunde modelleras utifrån area och den maximala lutningen med mycket hög förklaringsgrad (Figur 2). Hela 92% av variansen i log V kunde förklaras utifrån log A och  $\text{lut}_{max50}$ :

$$\log V = 0,75 + 1,06 * \log A + 0,056 * \text{lut}_{max50}; R^2=0,92; n=6130; p<0,0001 \quad (1)$$

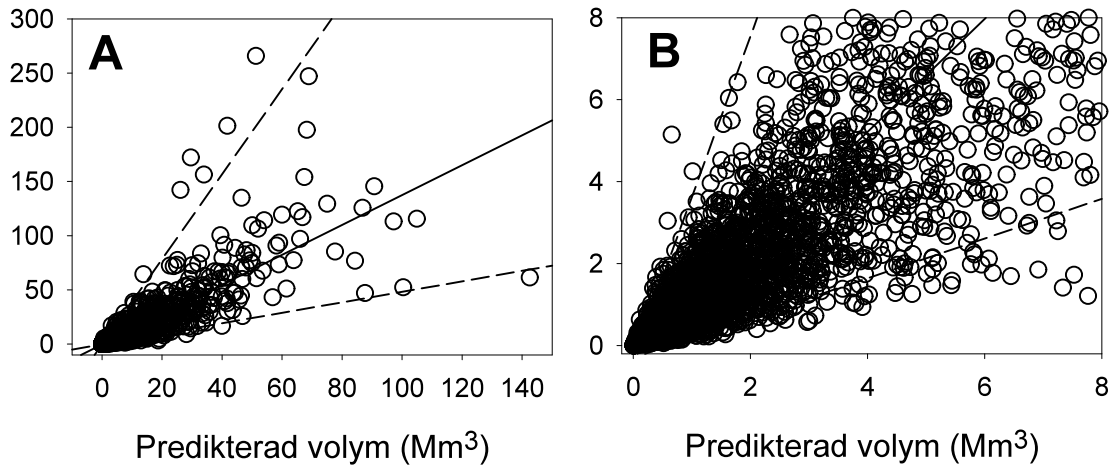
Strandlinjens flikighet, minimala lutningen och sjöomkrets, som framgick som måttligt betydelsefulla variabler i PLS-analysen, kunde antingen inte höja förklaringsgraden, eller var starkt interkorrelerade med antingen A eller  $\text{lut}_{max50}$ . Dessa variabler utelämnades därför från regressionsmodellen. Anmärkningsvärt var att sjöar i olika naturgeografiska regioner verkade följa samma samband; eller med andra ord, förklaringsgraden av log volym kunde inte höjas genom dela upp datamängden i naturgeografiska regioner. Möjligen gäller detta inte fjällsjöarna, som visade en signifikant större lutningskoefficient för log A än alla andra regioner ( $p<0,05$ ). Eftersom denna skillnad dock är baserad på bara 40 fjällsjöar, som dessutom är förhållandevis stora (median A,  $2,9 \text{ km}^2$ , att jämföra med median A för hela datamängden,  $0,21 \text{ km}^2$ ), måste skillnaden i lutningskoefficienten tolkas med försiktighet. Vi drar därför slutsatsen att sambandet mellan V, A och  $\text{lut}_{max50}$  gäller för alla svenska sjöar, med visst förbehåll för fjällsjöar.



**Figur 2.** Volym predikerad från sjöarea och maximala lutningen i en 50 m buffertzona från strandlinjen, mot observerad volym. Observera att volymen är log-transformerad. Den heldragna linjen visar följande regression:  $\log V = 0,75 + 1,06 * \log A + 0,056 * \text{lut}_{max50}$ ;  $R^2=0,92$ ;  $n=6130$ ;  $p<0,0001$ . De streckade linjerna visar prediktionsintervallet, dvs. intervallet, inom vilken predikerad log volym av en enstaka sjö ligger med 95% sannolikhet.

Ekvation (1) ska dock användas med försiktighet, eftersom den gäller log-transformerad volym. Den parameter som ska predikteras är dock volym, inte log-transformerad volym. Detta innebär att inte bara predikerad log volym måste omräknas till volym, utan även prediktionsintervallet för log volym, dvs. osäkerheten i prediktionen för en enstaka sjö. Figur 3 visar att det icke-log-transformerade prediktionsintervallet för volym är mycket stort. Den enda slutsatsen man med säkerhet kan dra från Figur 3 är att en sjö har en volym, så pass stor

är osäkerheten i prediktionen för en enstaka sjö. Trots att förklaringsgraden av logaritmen av volymen alltså är så hög som 92 %, så kan man ändå inte prediktera volymen av en enstaka sjö med någon vidare statistisk säkerhet.



**Figur 3.** Volym predikerad från sjöarea och maximala lutningen i en 50 m buffertzonen från strandlinjen, mot observerad volym. A visar hela datamängden, B visar en förstoring för mindre sjöar, som utgör >90% av datamängden. Observera att volymen inte är transformerad, till skillnad från Figur 2. Den heldragna linjen visar den linjära regressionen, och de streckade linjerna visar prediktionsintervallet, dvs. intervallet, inom vilken predikerad volym av en enstaka sjö ligger med 95 % sannolikhet.

Vi försökte flera sätt för att förbättra modellens prediktionsförmåga, såsom att hitta bättre transformationer, att använda latenta gradienter från PCA som variabler i multipla linjära regressioner, och att använda interaktionstermer mellan variabler som prediktorer, men utan framgång. Inte heller en uppdelning av datamängden i naturgeografiska regioner eller i olika storleksklasser kunde förbättra prediktionsförmågan. Vi försökte även att hitta osynliga grupperingar i datamängden genom clustering (vilket identifierar liknande observationer) och genom partitioning (vilket identifierar grupper av X-värden som bäst predikterar Y-värden), men även detta utan att höja prediktionsförmågan av sjövolym.

Observera dock att denna mycket begränsade prediktionsförmåga gäller enbart vid prediktion för en enstaka sjö. Så snart man ökar antal sjöar som det predikteras för, så minskar prediktionsintervallet. Detta eftersom prediktionsfelet är slumpvis fördelat, så att felet till viss del tar ut varandra om man predikterar volymen för flera sjöar åt gången. Predikterar man volym för 5 sjöar åt gången, minskar prediktionsintervallets storlek (jämför Figur 2) med 55%; vid prediktion för 10 sjöar åt gången med 68%; och vid prediktion för 100 sjöar åt gången med 89%. Detta innebär att prediktion av volym är möjlig med acceptabel statistisk säkerhet, om man predikterar för minst 5 sjöar åt gången. Därmed blir det möjligt att beräkna ett tillrinningsområdes totala sjövolym med hjälp av ekvation (1), om det innehåller minst 5 sjöar. En möjlig tillämpning är att prediktera totala sjövolymen uppströms miljöövervakningsstationer i vattendrag, eftersom uppströms sjövolym och vattnets uppehållstid skulle kunna förklara en del av de observerade skillnaderna i vattnets kemiska sammansättning, och i olika stationers variation i tiden.

## Referenser

- Algesten, G., Sobek, S., Bergström, A.-K., Ågren, A., Tranvik, L.J., and Jansson, M. 2004. Role of lakes for organic carbon cycling in the boreal zone. *Global Change Biology* **10**: 141-147.
- Engberg A. 2002. Produktspecification Svenska CORINE Marktäckedata, SCMD-0001, Lantmäteriet.
- Håkanson, L. 2004. Lakes: Form and Function. The Blackburn Press, Caldwell, New Jersey.
- Håkanson, L., and Peters, R.H. 1995. Predictive Limnology. SPB Academic Publishing, Amsterdam.
- Hamilton, S.K., and Lewis, W.M. 1990. Basin morphology in relation to chemical and ecological characteristics of lakes on the Orinoco river floodplain, Venezuela *Archiv Fur Hydrobiologie* **119**(4): 393-425.
- Hellström, T. 1991. The effect of resuspension on algal production in a shallow lake. *Hydrobiologia* **213**(3): 183-190.
- Jeppesen, E., Sondergaard, M., Jensen, J.P., Havens, K.E., Anneville, O., Carvalho, L., Coveney, M.F., Deneke, R., Dokulil, M.T., Foy, B., Gerdeaux, D., Hampton, S.E., Hilt, S., Kangur, K., Kohler, J., Lammens, E., Lauridsen, T.L., Manca, M., Miracle, M.R., Moss, B., Noges, P., Persson, G., Phillips, G., Portielje, R., Schelske, C.L., Straile, D., Tatrai, I., Willen, E., and Winder, M. 2005. Lake responses to reduced nutrient loading - an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology* **50**(10): 1747-1771.