

1 Analyse van de transitiematrix

In bijlagen 1 - 3 hebben we alle rekenstappen gebundeld samen met de R-code. Er is ook een rekenblad in Excel opgemaakt om vlot een overzicht te krijgen van de resultaten.

1.1 De basisgegevens

Tabel 1.1 vat de basisgegevens samen onder de vorm van een transitiematrix. Een transitiematrix is een kruistabel van het landgebruik van de eerste en tweede ronde. Een rij verwijst naar de toestand (bos of geen bos) tijdens de eerste cyclus ("timestep" t_1) en een kolom naar de toestand tijdens de tweede cyclus ("timestep" t_2). De cellen in de tabel geven het aantal rasterpunten in elke transitie categorie. De tabel geeft aan voor hoeveel rasterpunten de toestand gelijk blijft (de cellen op de diagonaal) en voor hoeveel de toestand verandert.

Voor sommige rasterpunten kon de toestand (het grondgebruik) niet bepaald worden omdat er geen toegang mogelijk was: de punten waren ontoegankelijk of de grondeigenaars weigerden toegang. Maar voor de meeste ontbrekende waarden is er wel informatie beschikbaar uit een andere ronde. Voor slechts 5 rasterpunten was er helemaal geen informatie beschikbaar. Van de 148 rasterpunten met een onbekend landgebruik op tijdstip 1, zijn 57 geen bos op tijdstip 2 en 86 bos op tijdstip 1. Van de 60 onbekende waarden op tijdstip 2, zijn er 30 geen bos en 25 wel bos op tijdstip 1.

Tabel 1.1 Transitie matrix. De rijen verwijzen naar de toestand op het eerste "tijdstip" (VBI-1), de kolommen naar de toestand op het tweede "tijdstip" (VBI-2). Samen definiëren een rij en een kolom een transitie categorie, t.t.z. de evolutie van de toestand op het eerste naar het tweede tijdstip: in 455 rasterpunten komt er bos bij (van 0 = "geen bos" naar 1 = "bos") en in 485 rasterpunten verdwijnt er bos (van 1 = "bos" naar 0 = "geen bos"). Er is permanent bos (1,1) in 2222 rasterpunten. Voor een aantal rasterpunten kon het landgebruik niet bepaald worden (NA = not available) omdat er geen toegang mogelijk was (ontoegankelijk of toegang geweigerd). Als we op basis van deze matrix "blindelings" de bosindex bepalen, dan is $BI_1 = 2732 / (27163 - 148) = 10,11 \%$ en $BI_2 = 2763 / (27163 - 60) = 10,19 \%$, met als trend $BI_2 - BI_1 = 0,08 \%$. Een kleine positieve trend dus. Na correctie zal blijken dat deze trend eigenlijk negatief is.

	$t_2: 0$ (geen bos)	$t_2: 1$ (bos)	$t_2: NA$ (onbekend)	som (t_1)
$t_1: 0$ (geen bos)	23798	455	30	24283
$t_1: 1$ (bos)	485	2222	25	2732
$t_1: NA$ (onbekend)	57	86	5	148
som (t_2)	24340	2763	60	27163

1.2 Hoe rekening houden met ontoegankelijke rasterpunten?

De ontbrekende rasterpunten weglaten zal leiden tot een onderschatting van de bosoppervlakte. De wegvallende rasterpunten zijn punten immers geselecteerd voor een terreinbezoek omdat ze meer dan gemiddeld kans maken om bos te zijn. Dat deze rasterpunten meer bebost zijn dan gemiddeld, kunnen we direct afleiden uit tabel 1.1. Van de 148 ontbrekende waarden bij VBI-1, is voor 143 punten de toestand in de tweede cyclus bekend, waarvan er 86 bos zijn, of 60,1 %. Dat is bijna zes keer zo hoog als de bosindex in Vlaanderen. Voor tijdstip 2 zijn er ook meer punten bos dan we uit de bosindex kunnen afleiden: van de 60 rasterpunten zonder informatie, is voor 55 de toestand bekend bij VBI-1 en hiervan zijn er 25 bos, of 45,5 %.

Als we voor deze punten de veronderstelling maken dat het landgebruik beperkt verandert, dan is het logisch om voor de ontbrekende waarden de uitkomst van de andere ronde over te nemen.

- VBI-1: van de 148 ontbrekende waarden bevinden er zich 86 punten (ongeveer 60 %) in bos in de tweede ronde. Als we deze punten als bos beschouwen voor VBI-1, dan verhoogt de bosindex met 0,32 % (4282 ha) voor VBI-1.
- VBI-2: van de 60 ontbrekende waarden waren er 25 punten (ongeveer 40 %) bos tijdens VBI-1. Deze als bos toevoegen verhoogt de bosindex met 0,09 % (1245 ha) voor VBI-2.

De correctie bij VBI-2 is veel kleiner dan bij VBI-1 om twee redenen. Er onderbreken bij VBI-2 minder waarnemingen, dus de correctie is haast per definitie kleiner. Daarenboven is het steekproefkader van beide meetcampagnes verschillend. Voor VBI-1 was het uitgangspunt de boskartering van 1990. Dat is een betrouwbaarder referentiepunt dan de orthofoto's van VBI-2 waar ook twijfelgevallen in zijn opgenomen om zeker geen bos te missen. Deze correctie is wellicht niet helemaal juist, maar zal niet tot grote fouten leiden tenzij de rasterpunten zonder toegang een veel grotere dynamiek zouden hebben dan de andere rasterpunten. Tabel 1.2 bevat de nieuwe transitietabel.

Tabel 1.2 De transitiematrix, gecorrigeerd voor de ontoegankelijke rasterpunten. Voor de ontoegankelijke rasterpunten nemen we het landgebruik over van de andere meetronde, waarvoor de toestand wel gekend is. We voegen dus $86 + 25 = 111$ punten toe aan de 2222 rasterpunten met permanent bos; $57 + 30 = 87$ punten worden toegekend aan de categorie 2x geen bos. Voor de bosindex (BI) zijn er twee scenario's uitgerekend. Met en zonder correctie. Bv. voor VBI-1: $(485 + 2222 + 25) / 27163 = 10,06 \%$ (de 25 ontbrekende waarden zijn wel bekend voor VBI-1) en $(485 + 2222 + 25 + 86) / 27158 = 10,38 \%$ (toevoegen van de 86 ontbrekende waarden, gegeven afkomstig van VBI-2). Het verschil is dus 0,32 %.

Transitiematrix	$t_2: 0$ (geen bos)	$t_2: 1$ (bos)		BI % (t_1)
$t_1: 0$ (geen bos)	23798 + 30 + 57 =	455	-30	
$t_1: 1$ (bos)	485	2222 + 86 + 25 =	-25	10,06 %
	-57	-86	-5	
BI % (t_2)		10,17 %		27163 - 5 =

1.3 Verschuivingen van het landgebruik

Tabel 1.2 is gedetailleerd om het effect van de correctie voor de ontoegankelijke punten te kunnen bespreken. Voor het vervolg gaan we verder met de finale transitietabel 1.3 waarbij de ontbrekende waarden vervangen zijn door de toestand op een ander ogenblik. De

transitietabel is de spil voor de berekeningen. Op basis van deze gegevens kunnen we alle andere kenmerken afleiden. Hier bekijken we i.h.b. de verschuivingen van het landgebruik.

De bosindex voor de eerste periode is de som $p_{10} + p_{11} = 10,38 \%$ en de bosindex voor de tweede periode is de som $p_{01} + p_{11} = 10,27 \%$. Het verschil is $(p_{01} + p_{11}) - (p_{10} + p_{11}) = p_{01} - p_{10} = -0,11 \%$. Uit deze afleiding volgt dat alleen de "discordante" cellen (rood in de tabel) nodig zijn om de trend te bereken: in 455 punten komt bos bij, in 485 punten verdwijnt bos. Netto verdwijnen er 30 punten en dat stemt overeen met een daling van de bosindex met $100 \times 30/27158 = 0,11 \%$, wat gelijk is aan het verschil tussen $10,27 \%$ (BI-vbi-1) en $10,38 \%$ (BI-vbi-2). De 2333 punten die permanent bos zijn, vallen tegen elkaar weg, bij het maken van het verschil.

Een ander punt is dat de bruto-verandering (het totaal aantal punten dat veranderd: $485 + 455 = 940 = 3,46 \%$) veel keer groter is dan de netto-verandering ($485 - 455 = 30 = 0,11 \%$). Een trend geeft dus geen goed beeld van de dynamiek van het grondgebruik. Bij de verwerking van de resultaten van de bosinventarisatie is het wenselijk om ook de bruto-verandering (verschuivingen in het landgebruik) te karakteriseren. De trend is eigenlijk maar het topje van de ijsberg.

Tabel 1.3 De transitiematrix (grijs) afgeleid uit tabel 1.1, waarbij de ontbrekende waarden ingevuld zijn met de gegevens van de meetcampagne waarvoor de toestand wel bekend is. De indices in $n_{00}, n_{01}, \dots$ verwijzen naar de toestand op t_1 (VBI-1) en t_2 (VBI-2): 0 = geen bos, 1 = wel bos; dus n_{01} = het aantal rasterpunten waar bos bijkomt & n_{10} = het aantal rasterpunten waar bos verdwijnt. n_{10} en n_{01} zijn de zogenaamde discordante meetparen, in tegenstelling tot de cellen op diagonaal waar de toestand niet veranderd is; i.h.b. is n_{11} het aantal rasterpunten met permanent bos. De $p_{00}, p_{01}, \dots$ zijn berekend t.o.v. de totale steekproefgrootte N (hieruit zijn de 5 rasterpunten weggelaten waarvoor alle informatie ontbreekt).

Transitiematrix	$t_2: 0$ (geen bos)	$t_2: 1$ (bos)	$BI \% (t_1)$
$t_1: 0$ (geen bos)	$n_{00} = 23885$ $p_{00} = 87,9 \%$	$n_{01} = 455$ $p_{01} = 1,68 \%$	
$t_1: 1$ (bos)	$n_{10} = 485$ $p_{10} = 1,79 \%$	$n_{11} = 2333$ $p_{11} = 8,59 \%$	$n_{t1} = n_{10} + n_{11} = 2818$ $p_{t1} = BI_1 = 10,38 \%$
$BI \% (t_2)$		$n_{t2} = n_{01} + n_{11} = 2788$ $p_{t2} = BI_2 = 10,27 \%$	$N = 27158$ $BI_2 - BI_1 = -0,11 \%$

1.4 De toets van McNemar voor symmetrie

Een geëigende, krachtige toets om na te gaan of bovenstaande transitie matrix symmetrisch (m.a.w. of $n_{01} \approx n_{10}$), is de toets van McNemar (Agresti, 2013):

$$Z_{McN} = \frac{n_{01} - n_{10}}{\sqrt{n_{01} + n_{10}}} \sim N(0,1)$$

Voor deze toets moeten we alleen informatie hebben over de cellen die veranderen (de zogenaamde *discordante* paren n_{01} & n_{10}).

Uit de transitie matrix (tabel 1.3) volgt:

$$Z_{McN} = \frac{455 - 485}{\sqrt{455 + 485}} = -0,98 \quad \text{met} \quad p_{McN} = 0,327$$

De (tweezijdige) p-waarde is groter dan 0.05 (significantieniveau 5 %) en - hiermee equivalent - de z-waarde is kleiner dan 1,96. We kunnen de nulhypothese dat er symmetrie is niet verwerpen.

Toetsen voor symmetrie komt eigenlijk neer op toetsen voor trend. We komen hier later op terug.

2 Evaluatie van de proefopzet

2.1 De correlatiecoëfficiënt

Doordat de VBI met een vast raster werkt, zijn de opeenvolgende metingen niet onafhankelijk van elkaar. Als een rasterpunt in bos valt, dan is de kans groot dat bij de volgende ronde, het rasterpunt in bos valt. Dat kunnen we aflezen uit de transitie matrix. Van de 2818 punten gelegen in bos bij VBI-1 zijn er nog 2333 punten in bos gelegen bij VBI-2 (82,8 %). Een manier om de verwantschap tussen de eerste en de tweede ronde uit te drukken is de correlatiecoëfficiënt ρ . Hier hebben we echter te maken met binaire variabelen (1 of 0 naargelang bos of niet), maar we kunnen hiervoor evengoed de correlatie berekenen als voor continue variabelen. Het is echter mogelijk een beknopte formule af te leiden, zodat we ρ rechtstreeks op basis van de transitie matrix kunnen berekenen (Agresti, 2013):

$$\rho = \frac{p_{11} - p_{t1} p_{t2}}{\sqrt{p_{t1} (1 - p_{t1}) p_{t2} (1 - p_{t2})}} = 0,813$$

Noot: om de notatie niet onnodig complex te maken, zijn de symbolen in bovenstaande formule als in tabel 1.3, maar in te vullen als proporties en niet als percentages (dus delen door 100 om de formule te gebruiken).

Per definitie ligt het bereik van de correlatiecoëfficiënt tussen twee grenzen: $-1 \leq \rho \leq 1$.

- een nul (het middelpunt van het bereik) stemt overeen met geen correlatie tussen de metingen (logisch);
- een negatieve waarde betekent dat er een tegengesteld verband is tussen de twee metingen (-1 verwijst naar een perfect maar tegengesteld verband);
- een positieve waarde dat er een gelijklopend verband is tussen de twee metingen (+1 komt neer op een perfect verband).

Zoals we voor het landgebruik kunnen verwachten, is ρ hoog (0,813).

2.2 Interpretatie

Een analyse van de formule voor ρ kan de betekenis verder verdiepen. Op het eerste zicht ziet de formule er complex uit, maar is in wezen eenvoudig. We ontrafelen de formule stapsgewijs:

- De noemer is een normeringsconstante die garandeert dat $-1 \leq \rho \leq 1$. Een bespreking hiervan zou te ver leiden en is hier verder weinig van belang. Van betekenis is vooral de teller. Voor de interpretatie, starten we met de tweede term.
- De tweede term in de teller is het product van de proportie bos zoals vastgesteld bij VBI-1 (p_{t1}) en de proportie bos bij VBI-2 (p_{t2}).
 - Het product van deze proporties is de proportie permanent bos die we theoretisch zouden verwachten indien tussen twee tijdstippen het landgebruik puur op basis van het toeval zou veranderen, net zoals we een dobbelsteen twee keer na elkaar zouden gooien zonder dat de eerste worp de tweede beïnvloedt.
 - De kans om twee keer na elkaar een zes te gooien is $\frac{1}{6} \times \frac{1}{6} = 1/36$. Indien we het experiment heel veel keer herhalen (theoretisch oneindig veel keer), dan zal 1/36 van alle worpen twee keer zes bevatten.
 - In het geval van de VBI "gooien" we 27000 keer, waarbij de kans op bos (een "zes") ongeveer 1/10 is (de bosindex is ongeveer 10 %, dus de kans dat een rasterpunt in bos terecht komt is 1/10). Dat impliceert dat de kans op twee keer bos na elkaar 1/100 is (meer precies, volgens tabel 1.3: $10,38/100 \times 10,27/100 = 0.0107$).
 - Moest het landgebruik veranderen zoals het gooien van een dobbelsteen, dan zou slechts 1 % permanent bos zijn bij een bosindex van 10 %. Dat is de nulhypothese.
- De eerste term drukt uit wat we feitelijk waarnemen, namelijk p_{11} , de proportie van de rasterpunten waar we twee keer bos vaststellen.
 - Volgens de transitie matrix in tabel 1.3 is 8,59 % (als proportie: 0,0859).
 - Het verschil $p_{11} - p_{t1}p_{t2} = 0,0859 - 0,0107 = 0,0752$ geeft de discrepantie tussen wat we waarnemen en wat voorspeld in de veronderstelling dat er geen verband is.
 - Het verschil is bijgevolg een maat die uitdrukt hoe sterk de verwantschap is tussen twee opeenvolgende metingen op hetzelfde rasterpunt.
 - Na deling door de noemer (de normeringsconstante) bekomen we $\rho = 0.813$; twee opeenvolgende waarnemingen op hetzelfde punt zijn bijgevolg sterk gecorreleerd.

2.3 Verband met de proportie discordante rasterpunten

Na in principe eenvoudige maar vrij omslachtige boekhoudkundige bewerkingen, kunnen we een heel direct verband aantonen tussen ρ en de proportie discordante rasterpunten:

$$\rho = \frac{1}{2} \frac{p_{t1}(1 - p_{t2}) + (1 - p_{t1})p_{t2}}{\sqrt{p_{t1}(1 - p_{t1})p_{t2}(1 - p_{t2})}} - \frac{1}{2} \frac{p_{01} + p_{10}}{\sqrt{p_{t1}(1 - p_{t1})p_{t2}(1 - p_{t2})}}$$

Als we veronderstellen dat $p_{t1} \approx p_{t2}$ (zoals een verschuiving van 0,1038 naar 0,1027), dan zal de eerste term in goede benadering gelijk zijn aan 1. Dan vereenvoudigt bovenstaande formule tot:

$$\rho \approx 1 - \frac{1}{2} \frac{p_{01} + p_{10}}{\sqrt{p_{t1}(1 - p_{t1})p_{t2}(1 - p_{t2})}}$$

en na invulling van de waarden uit de transitie matrix:

$$\rho \approx 1 - 5.4 (p_{01} + p_{10})$$

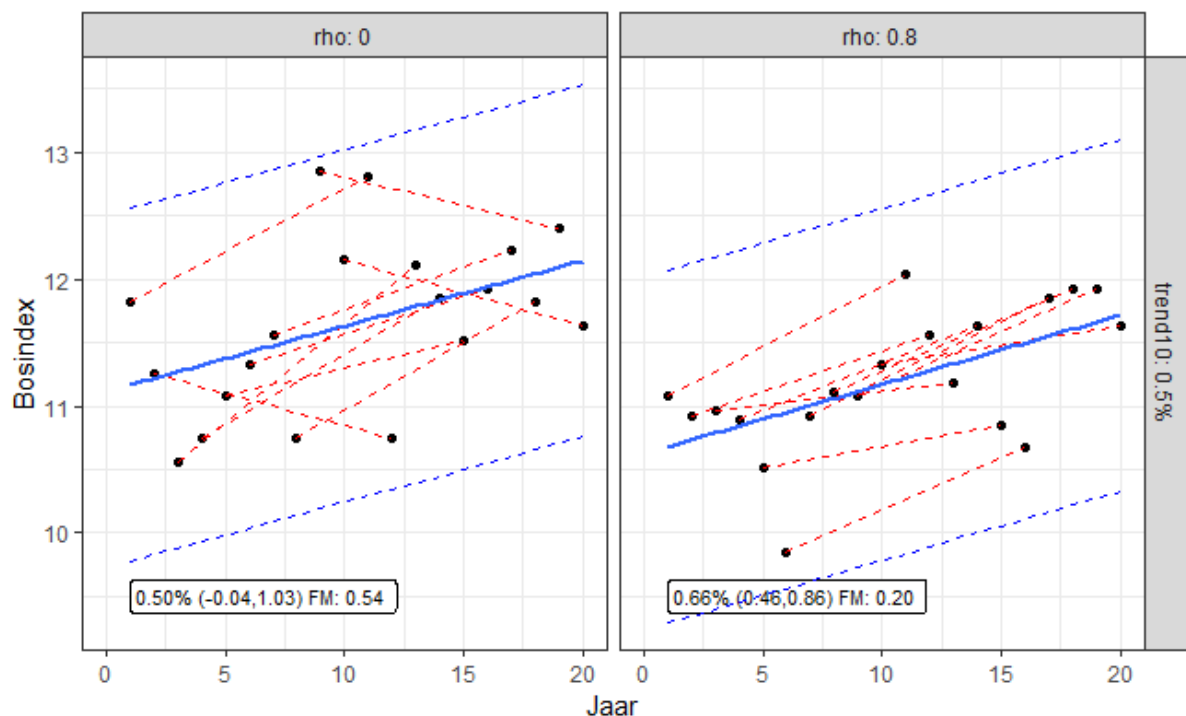
De correlatie is tegengesteld aan de proportie discordante punten. Het verband is zelfs lineair. De correlatie drukt de verwantschap uit tussen de gepaarde metingen; de proportie discordante paren verwijst naar de dynamiek, het aantal rasterpunten dat van toestand verandert. Een hoge dynamiek zal zich vertalen in een meer beperkte verwantschap. Dynamiek en correlatie zijn equivalente, maar tegengestelde grootheden.

2.4 Impact van de correlatie

De correlatie is heel belangrijk voor een bepaling van trend, maar ook in experimenten waar we twee behandelingen met elkaar vergelijken. Waar het in essentie over gaat, is dat we hetzelfde object of individu observeren in twee situaties. Door "paarsgewijs" te vergelijken, schakelen we verschillen tussen individuen uit. Het verschil is alleen te wijten aan het verschil in experimentele behandeling of het tijdsverschil. Hierdoor komt het effect beter in beeld. Hoe hoger de verwantschap tussen twee gepaarde metingen, des te groter de winst die we met een gepaarde proefopzet kunnen boeken. De statistische theorie kwantificeert deze winst.

Voor een gepaarde t-toets daalt de foutenmarge met een factor $\sqrt{1-\rho}$ (Sokal & Rolph, 2012). Voor $\rho = 0.8$ wordt FM een factor $\sqrt{5} = 2,24$ kleiner t.o.v. een proefopzet zonder gepaarde of gekoppelde meetpunten. We winnen meer dan een factor 2 door met een vast rooster te werken. Als we in elke ronde het raster vernieuwen, moeten we rekening houden met de steekproefvariabiliteit en hierdoor zal FM met een factor $\sqrt{5}$ verhogen. Ook figuur B1 geeft hier een illustratie van.

Een trendbepaling met gegevens van 2×5 jaar i.p.v. 2×10 jaar heeft hetzelfde effect omdat er geen koppeling meer is tussen de meetpunten. Daarbij komt dat we slechts de helft van de gegevens ter beschikking hebben, waardoor FM met een factor $\sqrt{2}$ stijgt. Een halvering van 20 (2×10) tot 10 (2×5) jaar verdubbelt daarenboven de FM met een factor 2. Het product van alle effecten leidt toe een verlies in nauwkeurigheid van $\sqrt{5} \times \sqrt{2} \times 2 = \sqrt{40} = 6.3$.



Figuur 2.1. Contrast tussen een onafhankelijke nieuwe steekproef en een vast rooster. Links zijn de rasterpunten niet vast, in de figuur rechts wel met een cyclus van tien jaar. De punten zijn de jaarlijkse schattingen van de BI op basis van 10 % (1/10) van de gemeten punten. Deze jaarlijkse schattingen zijn heel variabel omdat de steekproef klein is. **De rode stippellijnen** verbinden de jaarlijkse metingen met een tussenperiode van 10 jaar. Alleen rechts ontstaat hierdoor een helder

patroon die de trend weerspiegelt. In de figuur links zijn de rode lijnen heel variabel ten gevolg van de steekproefvariabiliteit.

De blauwe regressielijn is de best passende rechte voor de hele cyclus. Het kader onderaan vat de resultaten van de regressie samen: een schatting van de trend over een periode van tien jaar, het betrouwbaarheidsinterval en de foutenmarge (FM). In beide gevallen is de schatting van de trend goed, maar de foutenmarge rechts is meer dan een factor twee kleiner dan links. Hierdoor is de trend rechts significant en links niet.

2.5 De correlatiecoëfficiënt: berekeningen in R

Onderstaande R-code berekent de correlatie op twee manieren: een keer op basis van de formule en een keer via een omweg zodat we met de ingebouwde functie **cor** de correlatie kunnen berekenen (0,813) en met **cor.test** het bijhorende b.i.: (0,809; 0,817); *b.i.{ ρ }* is heel nauw want de steekproef is heel groot.

R-code

In R-base is er geen functie om rechtstreeks uit de transitietabel de correlatie te bepalen. Maar het is heel eenvoudig om zelf formule te programmeren. Een alternatieve werkwijze is uit transitiematrix eerst de binaire waarnemingen (1 of 0 naargelang bos of niet) te reconstrueren en hierop vervolgens de functie **cor** en **cor.test** toe te passen. De tweede variant is een omweg, maar is hier een illustratie dat we inderdaad hetzelfde resultaat bekomen. Een tweede voordeel is dat we het b.i. voor de correlatiecoëfficiënt kunnen berekenen met de functie **cor.test**.

```
## --- DATA STEP ---

## de gegevens (in een vector)
nTrans <- c(23885, 455, 485, 2333) ## aantallen
names(nTrans) <- c("n00", "n01", "n10", "n11")

## proporties (delen door het totaal aantal waarnemingen)
pTrans <- nTrans / sum(nTrans)
names(pTrans) <- c("p00", "p01", "p10", "p11")

## proportie bos
pt1 <- pTrans["p10"] + pTrans["p11"] ## proportie bos VBI-1
pt2 <- pTrans["p01"] + pTrans["p11"] ## proportie bos VBI-2
sprintf("%.2f", 100 * c(pt1, pt2)) ## print bosindex (x 100) met 2 dec.

## --- reconstructie binaire variabelen uit transitiematrix ---

## de aantallen in de transitiematrix stellen gepaarde metingen voor
## voor de 2x2 combinaties van bos/niet bos op de twee tijdstippen
## vb. n01 betekent dat voor dezelfde rasterpunten:
##   er n01 keer "0" (geen bos) was bij VBI-1 en,
##   er n01 keer "1" (wel bos) was voor VBI-2
## door voor alle combinaties deze reeks van 0 of 1 te creëren,
## reconstrueren we de oorspronkelijke binaire variabelen
## met de functie rep kunnen we deze herhalingen heel eenvoudig creëren
## noot: de lengte van deze variabelen is 27158; "length(vbi1_bin)"

vbi1_bin <- rep(c(0,0,1,1), nTrans) ## reconstructie voor VBI-1
## creëert: n00 x 0, n01 x 0, n10 x 1, n11 x 1
vbi2_bin <- rep(c(0,1,0,1), nTrans) ## reconstructie voor VBI-2
## creëert: n00 x 0, n01 x 1, n10 x 0, n11 x 1
## --- manuele berekening (toepassing formule) ---

vNum <- pTrans["p11"] - pt1*pt2 ## numerator
vDen <- sqrt(pt1 * (1 - pt1) * pt2 * (1 - pt2)) ## denominator
vRho <- as.numeric(vNum/vDen) ## as.numeric() verwijdert attributen
gRho <- 1/sqrt(1 - vRho) ## gain (winst door vast rooster)
c(rho = vRho, gain = gRho) ## resultaat: 0.813 & 2.313

## --- cor en cor.test (met gereconstrueerde variabelen) ---

cor(vbi1_bin, vbi2_bin) ## resultaat: 0.813 = identiek aan formule
cor.test(vbi1_bin, vbi2_bin) ## b.i.: (0.809, 0.817) = heel nauw
```

3 Analyse van de toestand

3.1 De schatting van de toestand: bosindex en bosareaal

Als n_{Bos} het aantal punten bos is (na correctie, zie hoger) op een totaal van N rasterpunten, dan geeft de proportie $p_{Bos} = n_{Bos}/N$ een schatting van bosindex (na vermenigvuldiging met 100).

$$BI [\%] = 100 \times p_{Bos} = 100 \times \frac{n_{Bos}}{N}$$

Vermenigvuldigen van de bosindex (in percentage) met de totale oppervlakte in Vlaanderen (13.522 km²) geeft het bosareaal in ha (omdat 1 % van 1 km² = 1 ha):

$$BA [ha] = BI [\%] \times 13522$$

3.2 Betrouwbaarheidsinterval voor de toestand

Volgens de statistische theorie is het aantal punten bos (n_{Bos}) binomiaal verdeeld met als parameters de werkelijke (maar onbekende) proportie bos P_{Bos} en steekproefgrootte N (Sokal & Rohlf, 2012).

Als zoals voor de VBI N groot is, dan is de steekproefproportie $p_{Bos} = n_{Bos}/N$ in goede benadering normaal verdeeld, met als verwachte waarde de werkelijke proportie P_{Bos} (hoofdletter P i.p.v. kleine letter p) en als variantie $P_{Bos}(1 - P_{Bos})/N$:

$$p_{Bos} \sim N(P_{Bos}, \frac{P_{Bos}(1 - P_{Bos})}{N})$$

Hieruit volgt voor het (95 %) betrouwbaarheidsinterval van de proportie (Sokal & Rohlf, 2012) (1,96 = 97.5 % percentiel van de standaardnormale verdeling):

$$b.i. \{p_{Bos}\} = p_{Bos} \pm FM_p$$

$$FM_p = 1.96 \times \sqrt{\frac{p_{Bos}(1 - p_{Bos})}{N}}$$

Vermenigvuldigen met 100 geeft de foutenmarge voor de bosindex en na vermenigvuldiging met de oppervlakte van Vlaanderen krijgen we de foutenmarge voor het bosareaal:

$$FM_{BI} (\%) = 100 \times FM_p$$

$$FM_{BA} (ha) = 13522 \times FM_{BI}$$

3.3 Berekeningen in R

Onderstaand kader geeft de R-code voor de berekeningen op basis van de functie **prop.test**. De resultaten zijn niet helemaal identiek aan tabel 1.4 omdat R een iets meer nauwkeurige, maar ook meer complexe berekening uitvoert. Met een grote steekproef, zijn de verschillen te verwaarlozen. We kiezen daarom de eenvoudiger formules die meer inzicht geven.

R-code: prop.test

De R-functie om betrouwbaarheidsintervallen voor proporties te berekenen is **prop.test**. Het eerste argument van de functie is het aantal rasterpunten bos, het tweede argument is het totaal aantal punten. Het totaal aantal rasterpunten bos in een bepaalde ronde (VBI-1 of VBI-2) is de som van het aantal punten met permanent bos (n_{11}) en het aantal punten dat verdwijnt (n_{10}) (voor VBI-1) of erbij komt (n_{01}) (voor VBI-2). Deze aantallen zijn te vinden in tabel 1.3.

```
(ptest_vbi1 <- prop.test(2333 + 485, 27158)) ## VBI-1: n11 + n10
(ptest_vbi2 <- prop.test(2333 + 455, 27158)) ## VBI-2: n11 + n01
```

De resultaten van `prop.test` "vangen we op" ze in een object. Deze werkwijze heeft als voordeel dat we resultaten verder kunnen bewerken: $\times 100$ om de proportie om te zetten naar percentages (%); $\times 13.522$ voor een omzetting naar het bosareaal (ha). We bundelen de resultaten in een vector, waardoor op meerdere getallen tegelijk de berekeningen mogelijk zijn.

```
## selecteer "estimate" en "conf.int" met "with()" uit de toets-objecten
## "diff(conf.int)/2" = helpt betrouwbaarheidsinterval = foutenmarge
## omzetting naar percentages  $\times 100$ 
```

```
BI_vbi1 <- with(ptest_vbi1, 100*c(estimate, diff(conf.int)/2, conf.int))
BI_vbi2 <- with(ptest_vbi2, 100*c(estimate, diff(conf.int)/2, conf.int))
```

```
## omzetting naar bosareaal
```

```
opp_V <- 13522 ## oppervlakte Vlaanderen in km²
BA_vbi1 <- BI_vbi1 * opp_V
BA_vbi2 <- BI_vbi2 * opp_V
```

```
## bundelen van alle resultaten in 1 tabel om uit te printen
## sprintf formateert de resultaten;
## > 2 cijfers na de komma voor BI
## > 0 cijfers na de komma voor BA
```

```
BI_vbi <- rbind(sprintf("%0.2f", BI_vbi1), sprintf("%0.2f", BI_vbi2))
dimnames(BI_vbi)[[2]] <- c("BI", "FM", "BI-lwr", "BI-upr")
```

```
BA_vbi <- rbind(sprintf("%0.0f", BA_vbi1), sprintf("%0.0f", BA_vbi2))
dimnames(BA_vbi)[[2]] <- c("B2", "FM", "BA-lwr", "BA-upr")
```

```
## Alles bundelen in een tabel (data.frame) en uitprinten
```

```
ptest_Table <- as.data.frame(cbind(BI_vbi, BA_vbi))
ptest_Table
```

```
##      BI    FM BI-lwr BI-upr      B2    FM BA-lwr BA-upr
## 1 10.38 0.36  10.02  10.75 140309 4929 135455 145314
## 2 10.27 0.36   9.91  10.63 138815 4906 133985 143797
```

4 Analyse van de trend

4.1 De toets van McNemar

De geëigende toets om van gecorreleerde binomiale variabelen na te gaan of er een trend is, is de McNemar toets (Agresti, 2013):

$$Z_{McN} = \frac{n_{01} - n_{10}}{\sqrt{n_{01} + n_{10}}} \sim N(0,1)$$

Hierbij is n_{01} het aantal rasterpunten waar bos bijkomt en n_{10} het aantal waar bos verdwijnt. Deze toets is alleen gebaseerd op cellen die veranderen (de zogenaamde *discordante* paren). De andere cellen uit de transitietabel geven geen informatie over de trend. Wat we eigenlijk onderzoeken met de toets is of de transitietabel symmetrisch is, m.a.w. of $n_{01} \approx n_{10}$ (Agresti, 2013).

De toetsstatistiek is (in goede benadering) standaardnormaal verdeeld; vandaar de keuze van het symbool z voor de toetsstatistiek. De p-waarde is af te leiden uit de standaardnormale distributie. Een waarde van z (in absolute waarde) groter dan 1,96 zijn significant op het 5%-niveau.

Toegepast op tabel 1.3:

$$Z_{McN} = \frac{455 - 485}{\sqrt{455 + 485}} = -0,98$$

$$p_{McN} = 0,327$$

De (tweezijdige) p-waarde is groter dan 0.05 (significantieniveau 5 %) en - hiermee equivalent - de z-waarde is kleiner dan 1,96. We kunnen de nulhypothese dat er geen trend is niet verwerpen. De waargenomen trend is (helemaal) niet significant.

In de literatuur (en ook in R) wordt vaak het kwadraat van de toetsstatistiek genomen. Aangezien het kwadraat van een normale verdeling, chi-kwadraat verdeeld is met 1 vrijheidsgraad (Sokal & Rohlf, 2012), is evenzeer een toetsing mogelijk. Beiden toetsen zijn equivalent, maar na kwadrateren kunnen we niet meer direct aflezen of het gaat om een dalende, dan wel stijgende trend.

$$Z_{McN}^2 = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \sim \chi_1^2$$

Berekeningen in R: mcnemar.test

De R-functie voor de McNemar-toets is **mcnemar.test**. De functie verwacht de transitiematrix als input. Zoals hierboven uitgelegd, is de output niet de z-waarde, maar wel het kwadraat ervan.

De functie **mcnemar.test** past een continuïteitscorrectie toe, hierdoor verschillen de waarden een beetje van de formules. Een continuïteitscorrectie is alleen nodig bij kleine steekproeven en heeft als doel om de distributie van de toetsstatistiek te verbeteren (lees meer normaal verdeeld te maken). Met "correct = FALSE" kan de default-waarde aangepast worden en wordt de correctie niet meer doorgevoerd.

```
## --- DATA STEP ---

## de gegevens (de transitiematrix in een vector)
nTrans <- c(23885, 455, 485, 2333) ## aantallen
names(nTrans) <- c("n00", "n01", "n10", "n11")

## omzetten naar matrix-formaat
mTrans <- matrix(nTrans, ncol = 2, byrow = TRUE)
dimnames(mTrans) <- list(c("t1:0", "t1:1"), c("t2:0", "t2:1"))

## --- Manuele berekening (met de formules) ---

## de z-statistiek

zMcN <- (nTrans["n01"] - nTrans["n10"]) /
  sqrt(nTrans["n01"] + nTrans["n10"])

## de p-waarde (tweezijdig, vandaar x 2)
## lower.tail = FALSE => overschrijdingskans

pMcN <- 2 * pnorm(abs(zMcN), lower.tail = FALSE)

## z^2 (chi-kwadraat verdeeld); automatisch tweezijdige test
## lower.tail = FALSE => overschrijdingskans
```

```

zMcn_2 <- zMcn^2
pMcn_2 <- pchisq(zMcn^2, 1, lower.tail = FALSE)

## alle resultaten gebundeld in een vector
## de p-waarden van beide testen zijn identiek

vMcn <- c(zMcn, pMcn, zMcn_2, pMcn_2)
names(vMcn) <- c("zMcn", "pMcn", "zMcn_2", "pMcn_2")

##      zMcn      pMcn zMcn_2 pMcn_2
## -0.978  0.328  0.957  0.328

## --- McNemar-functie in R: mcnemar.test ---

## zonder continuïteitscorrectie (correct = FALSE)

mcnemar.test(mTrans, correct = FALSE) ## input = matrix

## McNemar's Chi-squared test
## data:  mTrans
## McNemar's chi-squared = 1, df = 1, p-value = 0.3

## met continuïteitscorrectie (default-waarde)

mcnemar.test(mTrans)

## McNemar's Chi-squared test with continuity correction
## data:  mTrans
## McNemar's chi-squared = 0.9, df = 1, p-value = 0.3

```

4.2 Standaardfout voor de trend van gecorreleerde proporties

Om het betrouwbaarheidsinterval voor de trend te berekenen, gaan we er (opnieuw) vanuit dat de proporties in een goede benadering normaal verdeeld zijn (wegens de grote steekproef).

Het verschil van twee normaal verdeelde variabelen is opnieuw normaal verdeeld met als verwachte waarde het verschil tussen de verwachte waarden van de twee variabelen apart en met als variantie de som van de twee varianties apart.

Maar (!!!) aangezien het om gepaarde metingen gaat, moeten we de variantie V_{ind} (geldig indien de proporties onafhankelijk = *independent* waren) corrigeren met een factor $1 - \rho$:

$$(p_{Bos;2} - p_{Bos;1}) \sim N(p_{Bos;2} - p_{Bos;1}, (1 - \rho) \times V_{ind})$$

$$V_{ind} = \frac{P_{Bos;1} (1 - P_{Bos;1})}{N} + \frac{P_{Bos;2} (1 - P_{Bos;2})}{N}$$

De vierkantswortel uit de steekproefvariantie van het verschil geeft de standaardfout (standard error) op de trend:

$$s.e.\{p_{Bos;2} - p_{Bos;1}\} = \sqrt{1 - \rho} \sqrt{\frac{P_{Bos;1} (1 - P_{Bos;1})}{N} + \frac{P_{Bos;2} (1 - P_{Bos;2})}{N}}$$

Ten opzichte van statistisch onafhankelijke proporties is de standaardfout een factor $1 \div \sqrt{1-\rho}$ kleiner. Voor $\rho = 0.813$ is de correctiefactor 2,313. Door de gepaarde proefopzet met vaste rasterpunten winnen we meer dan een factor 2 in de precisie van de schatting van de trend. Voor slechts een heel beperkt aantal rasterpunten is er een verandering van het landgebruik, wat zich vertaalt in een hoge correlatie en een grote winst in precisie op de trend.

4.3 Gepaarde t-toets voor een trend van gecorreleerde variabelen

Als er in werkelijkheid geen trend is, dan geldt $p_{Bos;2} = p_{Bos;1}$, zodat de verwachte waarde van de normale distributie gelijk is aan nul.

We kunnen toetsen of de waargenomen trend ($p_{Bos;2} - p_{Bos;1}$) significant van nul verschilt door te vergelijken met de standaardfout. Dat komt neer op een gepaarde t-toets (Sokal & Rohlf, 2012).

$$Z_{paired} = \frac{p_{Bos;2} - p_{Bos;1}}{\sqrt{1-\rho} \sqrt{\frac{p_{Bos;1}(1-p_{Bos;1})}{N} + \frac{p_{Bos;2}(1-p_{Bos;2})}{N}}} \simeq Z_{McN}$$

Aangezien de steekproef heel groot is, is de toetsstatistiek standaardnormaal verdeeld; vandaar dat we "z" als symbool hebben gekozen. Bij de R-code zal blijken dat de gepaarde t-toets samenvalt met Z_{McN} , nog een reden om "z" als symbool te kiezen.

4.4 Betrouwbaarheidsinterval voor de trend van gecorreleerde proporties

Op basis van de normale benadering, kunnen we ook het (95 %) betrouwbaarheidsinterval afleiden voor het verschil in gecorreleerde proporties:

$$b.i. \{p_{Bos;2} - p_{Bos;1}\} = (p_{Bos;2} - p_{Bos;1}) \pm FM_{p-trend}$$

$$FM_{p-trend} = 1.96 \times \sqrt{1-\rho} \sqrt{\frac{p_{Bos;1}(1-p_{Bos;1})}{N} + \frac{p_{Bos;2}(1-p_{Bos;2})}{N}}$$

Vermenigvuldigen met 100 geeft de foutenmarge voor de bosindex; vervolgens vermenigvuldigen met de oppervlakte van Vlaanderen geeft de foutenmarge voor het bosareaal:

$$FM_{BI-trend} (\%) = 100 \times FM_{p-trend}$$

$$FM_{BA-trend} (ha) = 13.522 \times FM_{BI-trend}$$

Berekeningen in R: diffpropci.mp

R bevat in de basisuitvoering geen functies om met gecorreleerde proportie te werken. De functie **prop.test** kan alleen statistisch onafhankelijke proporties verwerken. Hierdoor worden de b.i. te breed ingeschat als de correlatie positief is.

Het R-package **PropCIs** is specifiek ontworpen om b.i. te berekenen voor gecorreleerde binomiale variabelen, conform een paper van Agresti (2005). We gebruiken de functie **diffpropci.mp** van het R-package.

Net zoals bij de toets van McNemar is alleen het aantal discordante paren van belang (d.w.z. alleen de rasterpunten waar het landgebruik veranderd is). Daarnaast is ook de totale steekproefgrootte (27158) nodig om de precisie te berekenen.

Ter vergelijking zullen we ook "manueel" op basis van de formules de berekeningen uitvoeren. Ten slotte proberen we ook de functie **prop.test** en tonen aan dat voor gecorreleerde proporties het resultaat fout is.

```
## --- data step ---

## oppervlakte Vlaanderen
opp_V <- 13522 ## voor conversie van BI (%) naar BA (ha)

## transitiematrix rij per rij
n00 <- 23885; n01 <- 455; n10 <- 485; n11 <- 2333
nTrans <- c(n00, n01, n10, n11)
names(nTrans) <- c("n00", "n01", "n10", "n11")

## totaal aantal rasterpunten
nTot <- sum(nTrans) ## 27158

## omzetting naar proporties
pTrans <- nTrans/nTot
names(pTrans) <- c("p00", "p01", "p10", "p11")

## --- diffpropci.mp (package: PropCIs) ---

## laden van het package (eerst installeren met install.packages)

require(PropCIs)

## we voeren de test diffpropci.mp uit
## en slaan het resultaat op in het object ptest_diff
## de haakjes () zorgen ervoor dat het resultaat geprint wordt

(ptest_diff <- diffpropci.mp(n10, n01, nTot, conf.level = 0.95))

## schatting (estimate) en b.i. (conf.int) samen in een vector plaatsen
## meteen bereken we de foutenmarge = helft c.i. (diff(conf.int)/2)
## vermenigvuldigen met 100 om alles procentueel uit te drukken

vtest_diff <- with(ptest_diff, c(estimate, diff(conf.int)/2, conf.int))
vtest_diff <- vtest_diff * 100
names(vtest_diff) <- c("trend", "FM", "lwr", "upr")

vtest_diff ## bosindex
## trend      FM      lwr      upr
## -0.110  0.221 -0.332  0.111

## omzetten naar ha

vtest_diff * opp_V ## bosareaal
## trend      FM      lwr      upr
## -1494  2993 -4487  1500
```

```

## --- manuele berekening ---

## proportie bos voor elke meetcyclus en trend (in %)
pt1 <- pTrans["p10"] + pTrans["p11"] ## proportie bos VBI-1
pt2 <- pTrans["p01"] + pTrans["p11"] ## proportie bos VBI-2

## trend bosindex
pTrend <- (pt2 - pt1) * 100 ## procentueel = verandering bosindex

## Noot: aangezien "p11" gemeenschappelijk is, is onderstaande ook oké
## hierbij wordt het verschil berekend van de discordante cellen
## pTrend <- (pTrans["p01"] - pTrans["p10"]) * 100

## de correlatie (rho)
vNum <- pTrans["p11"] - pt1*pt2 ## numerator
vDen <- sqrt(pt1 * (1 - pt1) * pt2 * (1 - pt2)) ## denominator
vRho <- vNum/vDen ## resultaat: 0.813

## standaardfout van de trend (standard error)
seTrend <- sqrt(pt1 * (1 - pt1) + pt2 * (1 - pt2))/sqrt(sum(nTrans))
seTrend <- seTrend * sqrt(1 - vRho) ## correctie voor correlatie
seTrend <- seTrend * 100 ## omzetten naar percentage zoals pTrend

## gepaarde t-toets
## we bekomen krak (!) dezelfde waarde als voor McNemar
tPaired <- as.numeric(pTrend/seTrend) ## -0.979
pval <- 2 * pnorm(abs(tPaired), lower.tail = FALSE) ## 0.328

## 95% - betrouwbaarheidsinterval voor BI
FM <- qnorm(1 - 0.05/2) * seTrend
biTrend <- c(pTrend, FM, pTrend + c(-1,1) * FM)
names(biTrend) <- c("Trend", "FM", "lwr", "upr")

## 95% - betrouwbaarheidsinterval voor BA
baTrend <- biTrend * opp_V
names(baTrend) <- c("Trend", "FM", "lwr", "upr")

## print resultaat
prn_biTrend <- sprintf("%0.2f %%", biTrend)
names(prn_biTrend) <- names(biTrend)
prn_baTrend <- sprintf("%0.0f ha", baTrend)
names(prn_baTrend) <- names(baTrend)
data.frame(BI = prn_biTrend, BA = prn_baTrend)

## --- foutieve analyse: prop.test ---

## prop.test berekent ook b.i. voor trends/verschillen tussen proporties
## de proporties moeten echter onafhankelijk zijn !!!
## voor positief gecorreleerde variabelen leidt prop.test tot
## te brede betrouwbaarheidsintervallen
## een correctie met sqrt(1 - ρ) is nodig

pv <- c(n11 + n10, n11 + n01)
nv <- rep(nTot, 2) ## aantal rasterpunten
ptest_vbi <- prop.test(pv,nv)

## selecteer estimate en conf.int uit het toets-object
## diff(conf.int)/2 = helft betrouwbaarheidsinterval = foutenmarge
## omzetting naar percentages x 100

BI_vbi <- with(ptest_vbi, c(diff(estimate), diff(conf.int)/2, conf.int))
BI_vbi <- 100 * BI_vbi

```

```

names(BI_vbi) <- c("Trend", "FM", "lwr", "upr")

data.frame(BI = BI_vbi, BA = BI_vbi * opp_V)

## FM is groter dan bij diffpropci.mp
## we proberen de correctie sqrt(1 - ρ) en, inderdaad,
## na correctie zijn FM en b.i. vergelijkbaar met diffpropci.mp

FMcorr <- BI_vbi["FM"] * sqrt(1 - vRho)
BIcorr <- c(FMcorr, BI_vbi["Trend"] - FMcorr, BI_vbi["Trend"] + FMcorr)
names(BIcorr) = c("FM.corr", "lwr.corr", "upr.corr")

BI_vbi <- c(BI_vbi, BIcorr)
BA_vbi <- BI_vbi*opp_V

data.frame(BI = round(BI_vbi, 2), BA = round(BA_vbi, 0))

```

5 Overzicht en bespreking van de resultaten

5.1 Overzichtstabellen

De transitietabel is de spil van de berekeningen waaruit alle andere kengetallen berekend kunnen worden. Daarom herhalen we hier in tabel 5.1 de transitiematrix na correctie voor de rasterpunten waarvoor geen toegang was. Om de ontbrekende waarden aan te vullen, hebben we aangenomen dat het landgebruik in die rasterpunten niet veranderd is. Tenzij deze punten heel atypisch zouden zijn, is deze correctie een verbetering omdat we anders de bosoppervlakte onderschatten.

Tabel 5.1. De (gecorrigeerde) transitiematrix (grijs) en toestand en trend van de bosindex. De n_{ij} en de p_{ij} stellen het aantal en percentage rasterpunten voor in elke transitie categorie, waarbij de eerste index naar de toestand op tijdstip 1 verwijst en de tweede index naar de toestand op tijdstip 2. De cellen op de diagonaal zijn rasterpunten die niet veranderen; de cellen niet op de diagonaal zijn de zogenaamde discordante paren waar het landgebruik veranderd is.

Transitiematrix	$t_2: 0$ (geen bos)	$t_2: 1$ (bos)	BI % (t_1)
$t_1: 0$ (geen bos)	$n_{00} = 23885$ $p_{00} = 87,9 \%$	$n_{01} = 455$ $p_{01} = 1,68 \%$	
$t_1: 1$ (bos)	$n_{10} = 485$ $p_{10} = 1,79 \%$	$n_{11} = 2333$ $p_{11} = 8,59 \%$	$n_{t1} = n_{10} + n_{11} = 2818$ $p_{t1} = BI_1 = 10,38 \%$
BI % (t_2)		$n_{t2} = n_{01} + n_{11} = 2788$ $p_{t2} = BI_2 = 10,27 \%$	$N = 27158$ $BI_2 - BI_1 = -0,11 \%$

Tabel 5.2 geeft een overzicht van de resultaten die we hierna bespreken. De kolommen VBI-1 en VBI-2 geven de toestand. Het verschil tussen beide geeft de trend.

Tabel 5.2. Overzicht van de resultaten. Schattingen van de bosindex (BI) en het bosareaal (BA) voor de eerste (VBI-1) en de tweede (VBI-2) ronde van de Vlaamse bosinventarisatie. De eerste lijn geeft de puntschatting $\pm$ de foutenmarge (FM); de tweede lijn het volledig uitgeschreven 95% - betrouwbaarheidsinterval (b.i.). De laatste kolom geeft de trend, het verschil tussen VBI-1 en VBI-2.

	VBI-1	VBI-2	Trend
Bosindex (BI)	10,38 $\pm$ 0.36 % (10,01 %; 10,74 %)	10,27 $\pm$ 0.36 % (9,90 %; 10,63 %)	-0,11 $\pm$ 0,22 % (-0.33 %; 0,11 %)
Bosareaal (BA)	140.309 $\pm$ 4905 ha (135.405 ha; 145.213 ha)	138.815 $\pm$ 4881 ha (133.934 ha; 143.696 ha)	-1.494 $\pm$ 2992 ha (-4.486 ha; 1498 ha)
Correlatie Toets McNemar			$\rho = 0,813$ (0,809; 0,817) $z = -0,98$ (p-waarde = 0,327)

5.2 Samenvatting

Op basis van de analyse kunnen we met een aantal eenvoudige kengetallen kunnen we de situatie samenvatten.

- Schattingen van het bosareaal zijn voor VBI-1: **140.309 ha (10,38 %)** en voor VBI-2: **138.815 ha (10,27 %)**; afgerond is de foutenmarge ongeveer **5.000 ha (0,36 %)**.
- De trend is klein: **een daling** van ongeveer **1.500 ha** (0.11 %); de foutenmarge op het verschil is ca. **3.000 ha** (0.22 %). Tussen de eerste en tweede cyclus liggen ongeveer 15 jaar (cyclus 1: 1997 - 1999, cyclus 2: 2009 - 2018) ¹ wat impliceert dat het bosareaal ongeveer **100 ha/jaar** (1500/15) gedaald is.
- Deze trend is niet significant: de trend is kleiner dan FM waardoor we de nulhypothese niet kunnen verwerpen. Het b.i. gaat van -4500 ha tot 1500 ha en omvat de 0-waarde. Toch zijn de resultaten informatief.
- Volgens de ondergrens kunnen we niet uitsluiten dat er in werkelijkheid een **een negatieve trend** is van 4500/15 = **300 ha/jaar**. Maar dat is het meest negatieve scenario. Een daling die nog groter is, kunnen we met een betrouwbaarheid van 95 % uitsluiten.
- Volgens het meest optimistische scenario (bovengrens van de b.i.) is er **een stijging** van 1500/15 = **100 ha/jaar**.

Hierbij passen nog twee aanvullingen:

- $FM_{trend} = 3000 \text{ ha} < FM_{toestand} = 5000 \text{ ha}$. De schatting op de trend is nauwkeuriger dan de schatting op de toestand. Dat komt omdat we met gepaarde metingen werken waardoor de gegevens gecorreleerd zijn; $\rho = 0,813$ en hieruit volgt dat we $FM_{toestand}$ met een factor $\sqrt{2(1-\rho)} = 0,611$ moeten vermenigvuldigen om FM_{trend} te bekomen. De verhouding tussen beide is inderdaad ongeveer 0,6.
- De toets van McNemar, speciaal ontworpen om gecorreleerde binomiale variabelen te analyseren, is equivalent met de gepaarde t-toets op gecorreleerde proporties. De toets is telkens $z = -0,98$ met een p-waarde = 0,328.

¹ Oorspronkelijk was de VBI opgezet als een periodieke inventarisatie: om de tien jaar in een korte periode (3 jaar) een volledige opmeting. Voor de tweede ronde is gekozen voor een continue inventarisatie waarbij elk jaar een tiende van de steekproef opgemeten wordt (Wouters, 2008).

5.3 Tot besluit, nog een caveat

VBI-1 en VBI-2 zijn niet helemaal met elkaar te vergelijken en deze methodologische verschillen bemoeilijken de interpretatie van de trend.

De eerste ronde (VBI-1) ging door in de periode 1997 - 1999 (Waterinckx *en Roelandt*, 2001). Voor de tweede ronde (VBI-2: 2009 - 2018) werd de methodiek gereviseerd (Wouters *et al.*, 2008). We geven een opsomming van de voornaamste verschilpunten.

- De meetcyclus. De eerste bosinventarisatie werd afgerond in een periode van 3 jaar. Voor de tweede ronde werd besloten om over te schakelen van een periodieke naar een continue inventarisatie met een jaarlijkse bemonstering van 1/10 van de steekproef. Deze ongelijke spreiding in de tijd bemoeilijkt de interpretatie van de trend. Voor de eenvoud hebben we hier verondersteld dat er een gemiddeld tijdsinterval is van 15 jaar.
- Het steekproefkader. VBI-1 is gebaseerd op de Boskartering 1978 - 1990 (Waterinckx, *et al.* 2001). Rasterpunten die geen bos waren volgens de Boskartering werden per definitie niet bezocht op het terrein, maar bij het terreinbezoek vielen er wel punten weg die geen bos (meer) waren². Dat kan tot een onderschatting leiden. Voor de selectie van de meetpunten werd voor de tweede Vlaamse Bosinventarisatie (VBI-2) werden orthofoto's gebruikt om in eerste fase te beoordelen of een punt bos was of niet. Op basis van deze screening werden punten al dan niet op het terrein bezocht om na te gaan of het effectief om bos gaat. Deze procedure laat in principe toe om nieuw bos te ontdekken.
- Overgangszones. Voor VBI-1 werden sommige beboste punten verschoven en/of niet weerhouden omdat het punt niet aan een aantal criteria voldeed. Voor de VBI-2 werden deze punten teruggelegd en is er bewust gekozen om grens- en overgangszones op te nemen en. Deze keuze kan een invloed hebben op de trend in de bosoppervlakte.

Deze verschillen zullen voor de volgende (derde) ronde veel minder spelen, omdat de methode nu volledig gestandaardiseerd is. Maar voor de huidige trendbepaling is het noodzakelijk bovenstaande elementen in het achterhoofd te houden.

² In dat opzicht is het logisch dat de schattingen van het bosareaal op basis van de VBI-1 lager (10,38 - 10,27 %) liggen dan de schatting van de boskartering (10,83 %).