

# BMP autoclustering: vooral een kwestie van doen!

## Vergelijking van clustermethodes

---

*Leo Schaap en Antje Ehrenburg*

Al vele jaren inventariseren vrijwilligers van de Vereniging voor Natuur- en Vogelbescherming Noordwijk (VNVN) broedvogels in enkele plots in het Zuid-Hollandse deel van de Amsterdamse Waterleidingduinen (AWD). Broedvogels inventariseren is een boeiende hobby en levert bovendien een bijdrage aan wetenschappelijk onderzoek. Sinds 2011 kan een groot deel van het uitwerken van het veldwerk door het autoclusterprogramma van SOVON worden overgenomen. Met het autoclusteren wordt de BMP-methode toegankelijker en komt het accent op het veldwerk te liggen. Dit artikel behandelt het BMP-monitoren en legt verschillen uit tussen hand- en autoclustering.

### Broedvogels monitoren

Om een broedvogelinventarisatie volgens het Broedvogel Monitoring Project (BMP) van SOVON uit te voeren zijn door SOVON regels opgesteld (van Dijk en Boele, 2011). Het veldwerk gebeurt volgens een vrij strak schema in een klein telgebied. Het telgebied moet in maart-juni minstens 7 (vogelarm gebied) tot 10 maal (vogelrijk gebied) volledig worden onderzocht, meestal omstreeks zonsopgang en minstens eenmaal 's nachts. Alle waargenomen soorten (inclusief exoten) worden met hun afgekorte naam en met speciale symbolen voor vijf typen waarnemingen zoals zang of nestindicerend gedrag genoteerd op de **veldkaart**.

De veldkaart is het belangrijkste stukje gereedschap om de inventarisatie uit te voeren. Het is essentieel om een goede kaart te gebruiken waarop zoveel mogelijk details te onderscheiden zijn. Ook een Google luchtfoto is goed te gebruiken en geeft verrassend veel details van de AWD. De waarnemingen moeten zo precies mogelijk op de kaart worden genoteerd. Dat valt niet altijd mee omdat er soms weinig ruimte is om alle waarnemingen te noteren. Daarbij komt dat het op gehoor bepalen van de plaats niet altijd nauwkeurig kan worden uitgevoerd. Je hoort een Nachtegaal in 'die richting' ongeveer in 'dat struikje' terwijl er ook over de eigen positie onzekerheid kan bestaan. Daarnaast is het van belang om de waarnemingen te onderscheiden conform de SOVON richtlijnen in: volwassen exemplaar in broedbiotoop, paar in broedbiotoop, ter-

ritorium indicierend gedrag, nest indicierend gedrag en nesten tellen. De meeste van deze waarnemingen worden als 'geldige waarneming' gezien. Voor het duinterrein is vooral het territorium indicerende gedrag zoals de zang, de meest voorkomende geldige waarneming.



*Boompieper, broedvogel in het Hoekgatterduin. Foto: George Hageman*

Na de inventarisatierondes worden per soort alle geldige waarnemingen op één kaart genoteerd, de zogenaamde **soortkaart**. Wat een geldige waarneming is wordt aangegeven in de BMP-handleiding en de bijbehorende tabel. Daarna worden op de soortkaart clusters van waarnemingen gevormd op basis van regels en voorbeelden die in de BMP-handleiding zijn opgenomen:

- Betrek zoveel mogelijk waarnemingen van verschillende bezoeken in een cluster.
- Een cluster mag maar één waarneming per bezoek bevatten.
- Een waarneming mag niet verder dan de fusiegrens van het uitgangscuster liggen.
- Een cluster mag maar 1,5 maal de fusiegrens groot zijn.
- Een cluster moet minimaal één waarneming binnen de datumgrenzen hebben.
- Een cluster moet voldoen aan een minimaal aantal geldige waarnemingen gebaseerd op een berekening van het aantal normbezoeken.

- Een cluster gebaseerd op twee waarnemingen is alleen geldig wanneer de waarnemingen minimaal 10 dagen na elkaar zijn gedaan.

Op basis van deze regels dienen zo veel mogelijk waarnemingen samengevoegd te worden tot een cluster, zodat zo weinig mogelijk clusters worden gemaakt. In de handleiding worden voor het vormen van clusters voorbeelden gegeven. Een cluster geldt als een territorium als aan de genoemde criteria wordt voldaan. Vooral omdat de BMP-methode soortspecifiek is - iedere soort heeft zijn eigen regels - is het handmatig uitwerken van de soortkaart erg arbeidsintensief en voor veel waarnemers bepaald geen geliefd onderdeel van het inventariseren.

## BMP Hoegatterduin 2010

Sinds 2011 biedt SOVON de mogelijkheid om na het online invoeren van de waarnemingen deze automatisch te laten clusteren. Om te onderzoeken of er verschillen zijn met handmatig clusteren heb ik van het jaar 2010 zowel een handclustering als een autoclustering gemaakt. Er is gekozen voor vergelijking van 2010 omdat de SOVON-handleiding met ingang van 2011 is aangepast. De jaren vóór 2011 zijn dus beter te vergelijken, alle gebaseerd op de handleiding van 2004 (van Dijk, 2004).

In tabel 1 staan de resultaten van hand- en autoclustering van het BMP-plot Hoegatterduin. Het valt direct op dat de autoclustering niet tot een trendbreuk leidt, althans, het kan op basis van deze cijfers niet worden aangetoond. De resultaten passen goed in het beeld van voorgaande jaren. Toch levert de vergelijking een aantal opmerkelijke verschillen op. Hoewel het totaal van de autoclustering 2010, 13 stuks (6,3%) territoria meer oplevert is het verschil in absolute zin, zonder plussen en minnen tegen elkaar weg te strepen, een stuk groter namelijk 29 stuks (14,1%). Dit is een ongecorrigeerd resultaat en ligt daarmee in dezelfde orde van grootte van elders in de AWD uitgevoerd onderzoek met het autoclusterprogramma van Van der Goes en Groot (van Groen, Oosterbaan en Ehrenburg, 2011). Dit onderzoek is professioneel uitgevoerd en leidde na onafhankelijke controle tot territoriumverschillen van 7, 13 en 4 procent voor drie verschillende plots. Een ander onderzoek van de VNVN uitgevoerd door Jaap Eisenga en Koene Vegter met gegevens van het BMP-plot in de AWD Wolfsveld-West 2011 levert ongecorrigeerd een vergelijkbaar verschil op van circa 15%.

*Tabel 1: Resultaten van clustering 2006 – 2011*

<b>BMP Hoegatterduin 2006 - 2011</b>								
Methode	Hand	Hand	Hand	Hand	Hand	Auto	Auto	
Jaar	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2010</b>	<b>2011</b>	
Totaal aantal territoria	157	181	207	187	193	206	195	
Aantal soorten	30	31	37	34	37	36	33	

De verschillen zijn per soort onderzocht en daarvan wordt in het periodiek Natuuronderzoek van de AWD verslag gedaan (Schaap, 2012). Uit die analyse blijkt dat het merendeel van de verschillen te maken heeft met het verkeerd toepassen van de handmatige clustering. Na correctie hiervan blijven een aantal afwijkende autoclusters over. Het aantal fout geïnterpreteerde territoria wordt door autoclustering aanzienlijk verkleind tot minder dan 5% (9 gevallen) waarover discussie blijft bestaan. Dit wordt geïllustreerd aan de hand van een paar voorbeelden.

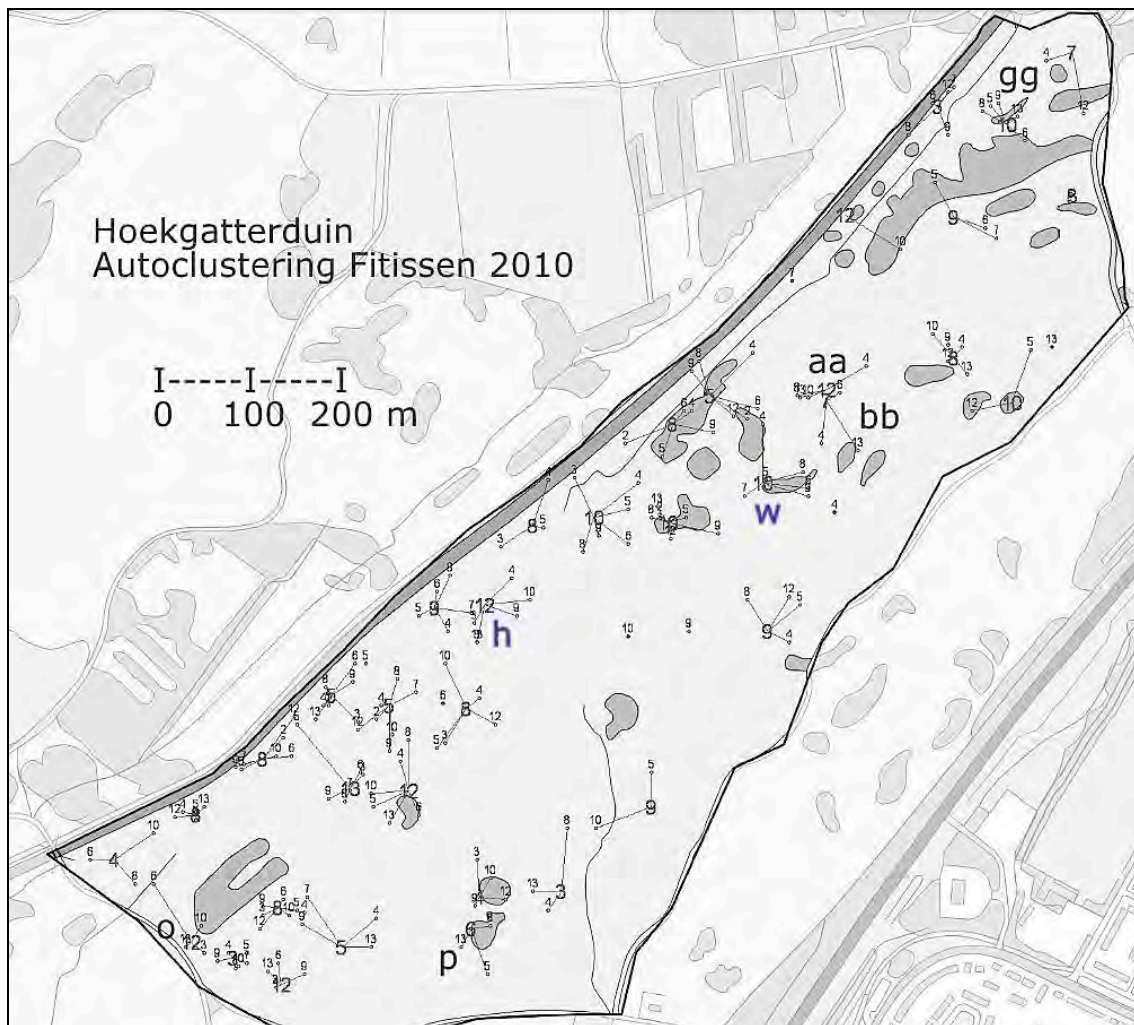
## Fitis

In het onderstaande voorbeeld wordt aandacht geschonken aan de autoclustering van de Fitis die in 2010 in het Hoekgatterduin uitkwam op 36 territoria. Handmatig waren dat er echter maar 30. Wat is hier aan de hand?

Als alle waarnemingen in een tabel netjes naast elkaar worden gezet blijkt dat autoclustering 4 territoria teveel heeft berekend. Deze clusters zijn gemerkt met **o**, **p**, **bb** en **gg** (dit heeft te maken met de analyse die is uitgevoerd). Deze clusters kunnen naadloos overgaan in naburige clusters omdat ze binnen de fusiegrens van 100 meter liggen en hun waarnemingen niet in nabijgelegen clusters voorkomen.



*Fitis, talrijkste broedvogel Hoekgatterduin 2010. Foto: George Hageman*



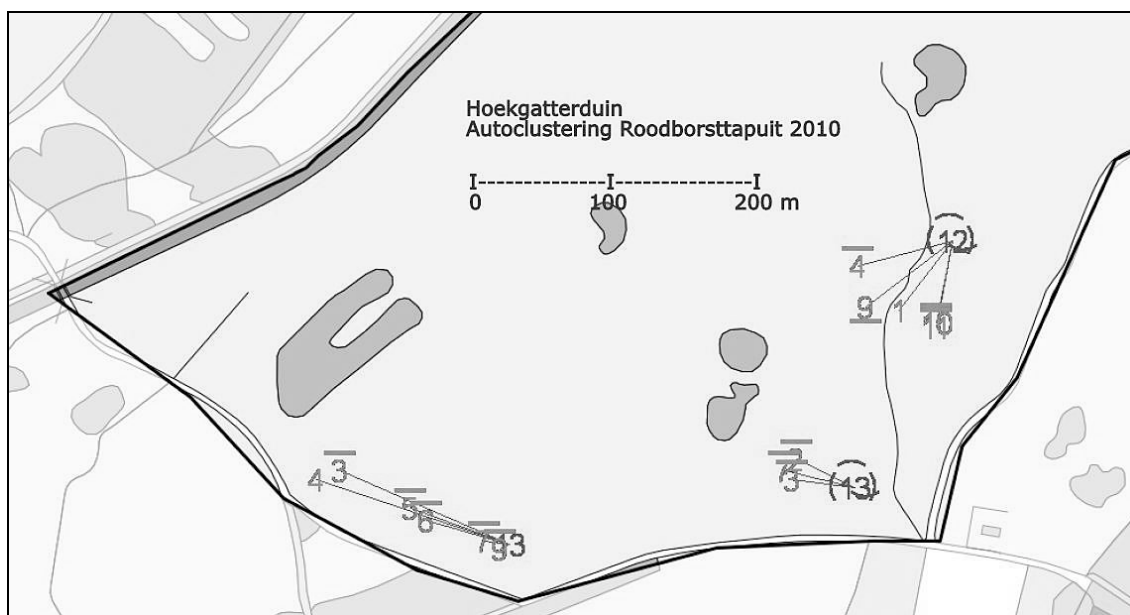
Figuur 1. Autoclustering Fitis 2010 Hoekgatterduin

Daarnaast blijkt dat tijdens het handmatig clusteren twee clusters over het hoofd zijn gezien, in de buurt van **h** en **w**. Dit komt omdat er zoveel handmatige waarnemingen op een hoopje liggen dat geen goed onderscheid kon worden gemaakt. Na een nauwkeurigere analyse en correctie blijven er 32 territoria over. Autoclustering leverde dus 4 territoria teveel en heeft waarschijnlijk te maken met onvolkomenheden in de software.

## Roodborsttapuit

Al jaren zijn in het Hoekgatterduin twee paartjes Roodborsttapuiten aanwezig die ieder jaar hetzelfde territorium bezetten. Dit bleek met de autoclustering niet het geval te zijn: er werden drie in plaats van twee territoria berekend. Hier is iets bijzonders gebeurd dat aan de hand van de clusterkaart wordt uitgelegd (figuur 2). De kaart ziet er iets anders

uit dan die van figuur 1, het is een keuze die in het programma kan worden gemaakt. In figuur 2 zijn de verschillende typen geldige waarnemingen te zien: die met een cirkel zijn nest indicierend,  $\bar{X}$  betekent de waarneming van een paartje,  $X$  betekent territorium indicerende waarneming zoals zang, en  $\bar{X}$  betekent een volwassen exemplaar. Tijdens het 12e en 13e bezoek zijn juveniele Roodborsttapuiten gezien die nog in het territorium rondscharrelden onder het toezien oog van de ouders.

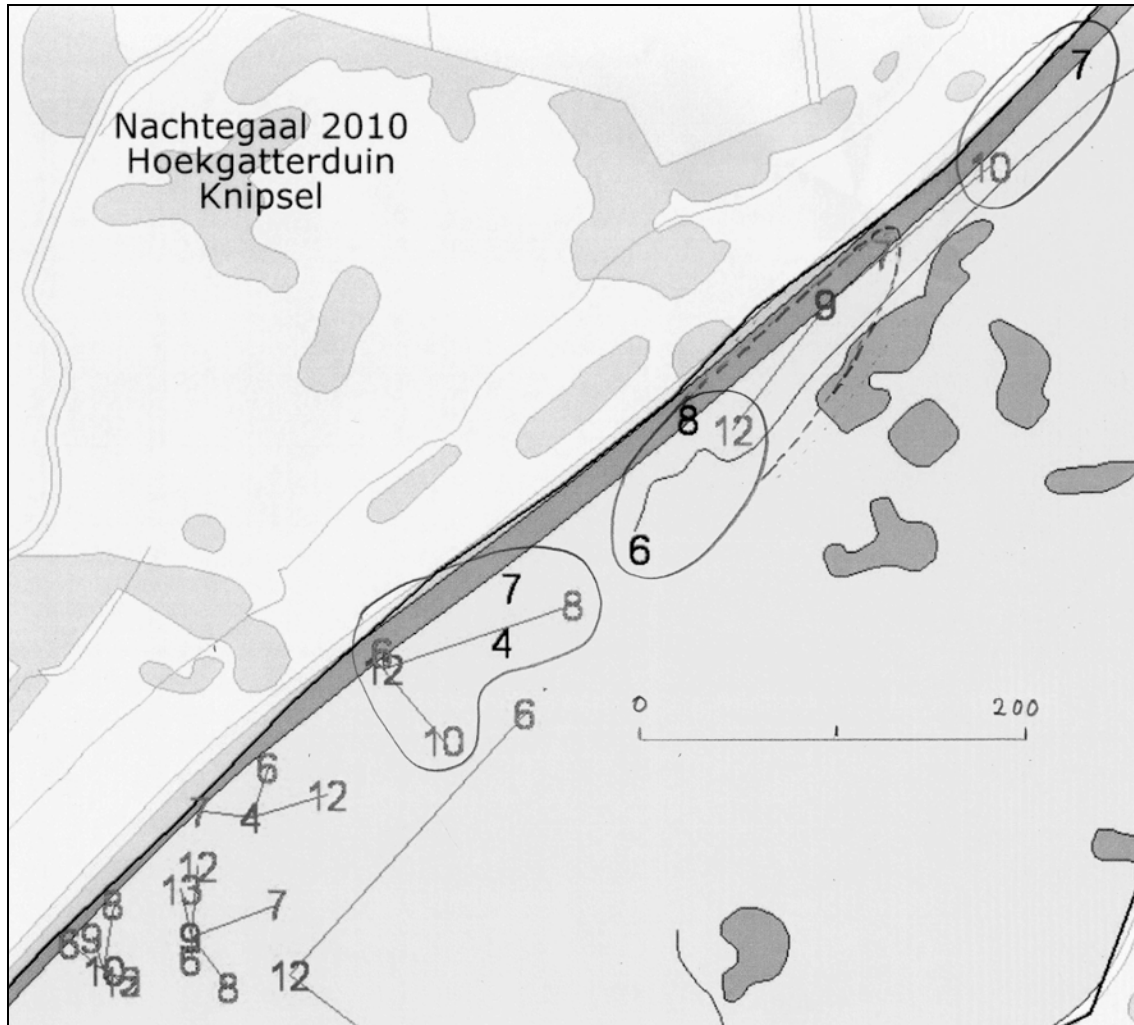


*Figuur 2. Autoclustering Roodborsttapuit Hoegatterduin 2010. Cijfers geven waarnemingen tijdens bezoekdata weer (1 = waargenomen bij bezoek 1, etc.). De cijfers waar lijntjes samenkomen geven de door de computer bepaalde uitgangswaarneming aan (12, 13 en 13). Samen met door lijntjes verbonden waarnemingen leidt dit tot zogenaamde autoclusters.*

Volgens de voorschriften is iedere nest-indicerende waarneming een apart territorium, terwijl het in dit geval toch echt om dezelfde juvenielen gaat. De waarneming van het 13<sup>e</sup> bezoek had dus niet tot een apart territorium moeten leiden. Hoewel krap, vallen alle waarnemingen binnen de fusiegrens van 200 meter en de gecombineerde waarnemingen voldoen ook aan het criterium dat de cluster kleiner moet zijn dan 1,5x de fusiegrens.

## Nachtegaal

De Nachtegaal broedt in het Hoekgatterduin vooral in het duindoornstruweel langs het Oosterkanaal. In figuur 3 wordt een deel van de soortkaart weergegeven.



*Figuur 3. Deel van de soortkaart Nachtegaal Hoekgatterduin 2010. De lijnen tussen waarnemingen geven autoclusters aan, de 'ellipsen' geven de handmatige clusters weer. De bezoeken 9 t/m 13 liggen binnen datumgrens. De losstaande cijfers (zonder lijntje) geven de waarnemingen aan die bij autoclustering ongebruikt blijven.*

De handclustering levert 14 territoria op, via autoclustering 13. Het verschil is te wijten aan de waarnemingen rond autocluster rond waarneming 9 in het midden van het kaartje waarvan in de buurt liggende waarnemingen 6 en 8 (ten zuiden) en 7 (ten noorden) bij de autoclustering niet worden gebruikt. Als je dit handmatig wel doet (zoals voorgeschreven door de BMP-regels: zo veel mogelijk waarnemingen in een

cluster zien onder te brengen) en 6, 8 en 12 vormen een uitgangscuster, dan ligt 10 te ver weg zodat deze 10 samen met 7 een eigen territorium kan vormen. Het gaat er dus om wat je als uitgangscuster neemt en dat de regel 'cluster niet groter dan 1,5x fusiegrens' wordt toegepast. Kennelijk gebeurt dit automatisch niet altijd goed.

Verder valt op dat bij autoclustering bezoek 4, een bezoek buiten de datumgrenzen (!), als centrum van een cluster wordt gebruikt. Dit komt vaker voor, het is ook op de soortkaart van de Fitis te zien, en zou je handmatig niet doen. Clusteren doe je meestal op basis van het bezoek met de meeste waarnemingen binnen de datumgrenzen. Ook valt op dat er waarnemingen (zwart van kleur) niet gebruikt worden terwijl ze gemakkelijk aan een nabijgelegen cluster kunnen worden toegevoegd.

## Conclusies

Handmatige clustering leidt door de complexiteit van BMP gemakkelijk tot fouten. Autoclustering daarentegen is niet alleen uitgerust met een gebruiksvriendelijke interface, maar maakt ook de invoer van waarnemingen gemakkelijk en overzichtelijk. De administratie wordt goed en toegankelijk bijgehouden en betekent minder werk achteraf.

Het grootste voordeel is dat minder menselijke fouten optreden. In het geanalyseerde geval blijft een afwijking bestaan tot minder dan 5% van het totaal door autoclustering bepaalde aantal (Schaap, 2012). Bovendien zorgt autoclustering voor een betere standaardisatie die van groot belang is voor het vogelonderzoek.

Zoals hierboven is aangetoond is autoclustering niet op alle fronten ideaal. De autoclustering gaat anders dan handmatig zou gebeuren. Ecologische aspecten worden niet meegenomen, terwijl ook de autoclustermethode voor fouten zorgt. Handmatig wordt uitgaan van een **uitgangscuster** waar vanuit omliggende waarnemingen al of niet worden toegevoegd. Autoclustering daarentegen neemt een willekeurige waarneming als **uitgangswaarneming**, soms ligt deze waarneming zelfs buiten de datumgrenzen.

De waarnemer die overstapt naar autoclustering dient zich bewust te zijn van een paar verschillen omdat de handmatige spelregels niet helemaal gelijk zijn aan die van de computer:

- 1) Autoclustering werkt niet met een uitgangscuster maar met een uitgangswaarneming die zo te zien willekeurig wordt gekozen. Bovendien beperkt autoclustering de clustergrootte en komt nooit tot 1,5x de fusieafstand.
- 2) Autoclustering haalt soms een waarneming van ver terwijl er een andere waarneming van dezelfde ronde dichtbij ligt. Dat ziet er in verge-



lijking met een handmatige clustering heel raar uit. Digitaal maakt dat niet uit, het algoritme van de computer is wiskundig gericht en ecologische kenmerken komen er niet in voor.

- 3) Er is er geen mogelijkheid om bij autoclustering een 100% uitsluitende waarneming aan te geven. De autoclustering gaat er van uit dat alle waarnemingen 100% uitsluitend zijn, terwijl dat voor de handmatige methode anders is. SOVON benadrukt dat in het veld de goede afweging gemaakt moet worden: gaat het om een nieuw individu, of is het mogelijk dezelfde vogel die al eerder is genoteerd?
- 4) Er is met autoclustering geen mogelijkheid om gebruik te maken van grensgevallen. Een waarneming valt er binnen of er buiten. De administratie houdt het keurig bij maar ze worden niet gebruikt bij het berekenen van clusters. Een zekere manipulatie is natuurlijk mogelijk door deze waarnemingen binnen de gebiedsgrenzen te schuiven. Dat hier niet teveel gebruik van moet worden gemaakt lijkt duidelijk. Standaardisatie van een meting treedt pas op met zeer strikte regels ook al houdt een vogel zich niet aan gebiedsgrenzen.



*Roodborsttapuit: 2 broedparen in Hoekgatterduin. Foto: George Hageman*

Ten slotte nog een paar andere opmerkingen:

- Men dient er rekening mee te houden dat autoclustering zich niet handmatig laat corrigeren of voor een deel met de hand is in te vullen; het is óf het een óf het ander. Correctie achteraf in de SOVON database is mogelijk omdat de BMP-tabel op internet nog steeds handmatig kan worden ingevuld. Men kan dus voor deze optie kiezen.
- Autoclustering legt de territoriumstip regelmatig op een waarneming van vóór de datumgrenzen, iets wat je handmatig nooit zou doen. Uitgangspunt van handclustering zijn de waarnemingen van het bezoek met de meeste waarnemingen na de datumgrens.

Dit alles laat onverlet dat er grote waardering is voor het werk van SOVON. Het is een geweldige prestatie om een dergelijk groot automatiseringsproject in relatief korte tijd door te voeren. Daarbij is veel aandacht geschonken aan de gebruiksvriendelijkheid. Inventariseren wordt hiermee aantrekkelijker omdat het minder arbeidsintensief is geworden. Nu nog de mogelijkheid om de papieren veldkaart te vervangen door een tablet computer zoals een iPad. Dat zal nog een extra groot voordeel hebben: de geïntegreerde GPS-ontvanger zal de positie van de waarnemer exact weergeven, zodat daarover geen onzekerheid hoeft te bestaan. Daarnaast wordt, mits zorgvuldig aangepakt, de inventarisatie beperkt tot veldwerk, het ideaal van iedere vogelaar!

Ten aanzien van de autoclustermethode zal meer onderzoek gedaan moeten worden om tot een optimale methode te komen. Dit heeft vooral te maken met de grootte van de fusieafstand en het niet vormen van uitgangscusters rond de waarnemingen van het bezoek met het hoogste aantal binnen de datumgrenzen. In die zin volgt de autoclustermethode de handleiding niet. Daarnaast zou de invoer sterk verbeterd kunnen worden door een groter invoerscherm van de onderliggende luchtfoto. SOVON kan bovendien bijdragen aan een beter begrip door inhoudelijk meer en beter over autoclustering met vrijwilligers te communiceren om er uiteindelijk voor te zorgen dat hand- en autoclustering niet van elkaar gaan verschillen.

## Literatuur

- Dijk, A.J. van, 2004. Handleiding Broedvogel Monitoring Project, SOVON Vogelonderzoek Nederland, Beek-Ubbergen.
- Dijk, A.J. van, en Boele A. 2011. Handleiding SOVON Broedvogelonderzoek. SOVON Vogelonderzoek Nederland, Nijmegen.
- Groen, Frank van, Oosterbaan Bernhard en Ehrenburg Antje. 2011. Broedvogels clusteren met een computerprogramma, een uitkomst? In: Natuuronderzoek, natuurberichten uit de Amsterdamse Waterleidingduinen, Waternet, Jaargang 21, nummer 1, maart 2011.
- Schaap, Leo, 2012. SOVON auto- en handclustering vergeleken. In: Natuuronderzoek, natuurberichten uit de Amsterdamse Waterleidingduinen, Waternet, Jaargang 22, nummer 1, maart 2012.