

Automatische verwerking van gegevens suikerbietenonderzoek

Dr. A.C.P.M. van Swaaij en L.M. Withagen

Instituut voor Rationele Suikerproductie IRS

Postbus 32, 4600 AA Bergen op Zoom

telefoon (0164) 27 44 00, telefax (0164) 25 09 62

e-mail: vanswaaij@irs.nl

internet: www.irs.nl

Jaarlijks leveren proefvelden op vele locaties verdeeld over Nederland een aanzienlijk aantal monsters. Dat vraagt om een goede logistiek en registratie. Medewerkers van het IRS leveren jaarlijks via meerdere kanalen (o.a. handheldcomputers, weeginrichtingen en meetapparatuur) zo'n 300.000 afzonderlijke data van de proeven. Het verwerken van deze gegevens, maar ook het bepalen van de proefopzet is bij het IRS centraal georganiseerd. Het database-systeem dat hiervoor in gebruik is biedt zoveel ruimte, dat onderzoekers zelf gemakkelijk vanaf hun eigen werkplek data kunnen opvragen en gebruiken.

Trefwoorden: Database, Oracle, handheldcomputer, gegevensverwerking, proefveld

Het Instituut voor Rationele Suikerproductie (IRS) is sinds 1930 het onderzoeks- en kennisinstituut voor de suikerbietenenteelt in Nederland. De kern van het onderzoek vormen de proefvelden die worden aangelegd bij een groot aantal telers in Nederland. In het totaal gaat het momenteel om zo'n 150 proefvelden op ca. 75 locaties.

De rekenkamer van het IRS voert al vanaf het oprichtingsjaar centraal de proefopzet, verwerking, analyse en archivering van alle waarnemingsuitkomsten uit. Dit laatste heeft geresulteerd in een omvangrijk data-archief, waardoor o.a. dubblures in het onderzoek voorkomen kunnen worden. Daarnaast komt de centrale en uniforme verwerking de kwaliteit van het onderzoek ten goede.

Sinds 1970 kunnen de uitkomsten van wegenen en laboratoriumanalyses aan een computer gekoppeld worden. De verwerking van deze gegevens gebeurde in eerste instantie met in Basic geschreven programma's, later met Fortran-programma's.

In 1997 werd besloten om over te stappen naar een PC-netwerkomgeving. Tegelijkertijd leefde de wens om de verwerking en de analyse van de waarnemingsuitkomsten verder te automatiseren en een beter toeganke-

lijke database te creëren. Waar tot nu toe alles in eigen beheer was opgezet, werd hier de hulp van een extern informaticabedrijf ingeroepen.

Van proefopzet tot eindresultaat

Het proefveldseizoen (figuur 1) begint met de keuze van geschikte lokaties voor het onderzoek. De rekenkamer legt de gegevens over geschiedenis en eigenschappen van de gekozen proefvelden vast. In overleg met de onderzoeker maakt de statisticus een experimentele opzet voor ieder proefveld. Het schema van de proefopzet wordt afgedrukt en gaat mee naar het veld. Het moet bruikbaar zijn bij het zaaien, maar ook bij het uitvoeren van de verschillende behandelingen (bijvoorbeeld bemestingstrappen, ziekten- en onkruidbestrijding) en voor het doen van de veldwaarnemingen.

Waarnemingsuitkomsten komen gedurende het gehele teeltseizoen binnen en kunnen zeer gevarieerd in vorm zijn. Soms betreft het gemiddelde beoordelingscijfers van het gewas per veldje, andere keren zijn het tellingen (bijv. plantaantal) op een lager niveau van rijen binnen een veldje.

Vanaf september oogsten de proefveldmede-

werkers de bieten machinaal of met de hand. Zij bepalen de opbrengst per veldje en nemen afhankelijk van het type proef meerdere monsters. Deze worden in het IRS verder onderzocht op in- en externe kwaliteit.

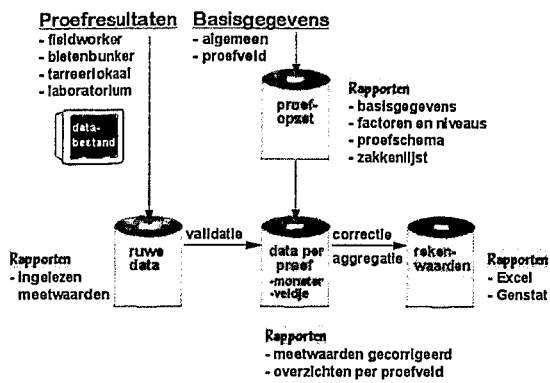
De uitkomsten van dat onderzoek aan zo'n 20.000 monsters per seizoen, alsmede de uitkomsten van de overige waarnemingen tijdens het oogsten gedaan, komen in de periode september - december samen in de rekenkamer. De meeste gegevens staan niet op zichzelf, maar zijn nodig om afgeleide waarden te berekenen. Een extra moeilijkheid bij de verwerking is dat niet alle waarden op het zelfde niveau zijn gemeten. Zo wordt de bruto-opbrengst op veldniveau bepaald, maar de verhouding bruto, netto, tarra op het niveau van submonsters van die veldjes.

De koppeling van alle gegevens aan de afzonderlijke objecten en de daaropvolgende berekeningen en statistische analyses moeten in zeer korte tijd worden uitgevoerd. Met name de verwerking van de uitkomsten van de rassenproeven staat onder grote tijdsdruk. Door de snelle verwerking kunnen telers al voor het nieuwe seizoen profiteren van de informatie over de nieuwe rassen.

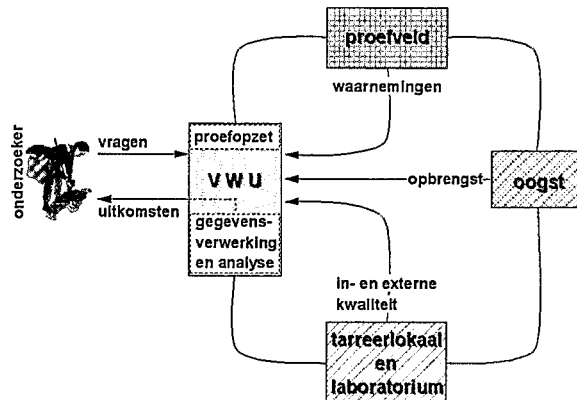
Aan het einde van het seizoen zorgen de medewerkers van de rekenkamer ervoor dat alle verzamelde waarnemingsuitkomsten op projectnummer gerangschikt in de zogenaamde proefveldboeken worden opgeslagen. Tenslotte worden de verantwoordelijke onderzoekers ondersteund bij de interpretatie en de presentatie van de uitkomsten.

Centrale Oracle database

In 1997 is het agro-informatiseringsbedrijf Q-Ray gevraagd om de gegevensverwerking om te zetten naar en verder te automatiseren in een PC-netwerk omgeving. Het uit-



Figuur 1 – De proefveldcyclus. VWU = database Verwerking-Waarnemings-Uitkomsten



Figuur 2 – Schematische weergave van in- en uitvoer bij de database VWU

gangspunt hierbij was het zoveel mogelijk beperken van handmatige invoer en een betere toegankelijkheid van de gegevens voor alle onderzoekers en ook voor later gebruik. Daarnaast moest het verwerkings-systeem voldoende flexibel zijn om allerlei verschillende datastructuren en de niveaus waarop ze worden verzameld aan te kunnen. Besloten werd met Oracle software een database applicatie op maat te maken en deze de naam VWU (Verwerking-Waarnemings-Uitkomsten) te geven.

Het principe van VWU is schematisch weergegeven in figuur 2. VWU wordt gevoed met de ruwe data van proefresultaten en met algemene en proefveldspecifieke gegevens. De ruwe data worden gekoppeld aan experimentele eenheden van de proeven en getoetst aan vooraf gedefinieerde meetwaarden (validatie) en vervolgens na correctie geaggregeerd naar proefveldniveau. Middels standaard aanwezige modules of via specifieke uitvoer (Excel, Genstat) kunnen onderzoekers de data bekijken.

Nauwkeurige voorbereiding

Elke proef is gedefinieerd voor een bepaald jaar en kent zijn eigen specifieke gegevens. Alle proeven vallen echter weer onder vaak langjarige projecten en subprojecten. Daarom zijn er ook veel gezamenlijke gegevens. Het invoeren hiervan gebeurt nu veelal slechts één keer, zodat bij het beschrijven van de proeven volstaan kan worden met kopiëren (bijvoorbeeld door aan te vinken) en eventueel aanpassen of

aanvullen van de betreffende basisgegevens. Het vullen van de VWU-database begint met het definiëren van de standaard basisgegevens. Deze basisgegevens bestaan uit:

- proefveldhouders en hun adresgegevens;
- onderzoeksprojecten en projectleiders;
- standaardfactoren en de daarbij behorende niveaus; bijvoorbeeld de factor ras met als niveaus de rasnamen;
- standaard meetvariabelen; hiervan zijn de meeteenheid, het niveau van de meting (rij, monster of het hoogste niveau: veldje), aantal decimalen en beperkingregels vastgelegd; tevens zijn hierin formules voor de berekening van afgeleide variabelen beschreven en is aangegeven hoe geaggregeerd moet worden naar een hoger niveau (bijvoorbeeld voor analytische gegevens van monsters een middeling en voor plant-aantallen in de rijen een optelling).

Het starten van een proef voor een bepaald jaar begint met het vastleggen van algemene gegevens over bijvoorbeeld de ligging en het bouwplan van het proefveld, de locatie (met de daarbij horende telergegevens) en de grootte van de veldjes in de proef. Daarna wordt aangegeven welke meetvariabelen en welke objecten (factoren en de bijbehorende niveaus) in de proef voorkomen. De invoer van deze specifieke proefveldgegevens is een stuk efficiënter door de beschikbaarheid van de standaard basisgegevens in oproepbare vensters. Het afvinken van keuzelijsten beperkt de handmatige invoer en de daarbij horende foutenkans.

Het proefveldschema is behalve van het aantal objecten tevens afhankelijk van het aantal herhalingen en de ligging van deze herhalingen in het proefveld. Dit wordt aan de database bekend gemaakt. Daarop genereert de database voor alle veldjes een volgnummer, beginnend links onder en eindigend rechts boven. Een ingebouwde lotingprocedure kent aan elk object binnen elke herhaling een random plaats toe (randomized-block-design). Het ligt in de bedoeling in de toekomst de database uit te breiden met andere lotingprocedures, zodat ook split-plot of incomplete-block-designs mogelijk worden. Na de loting is ieder veldje gekoppeld aan een herhaling (A,B,...) en aan een objectnummer (1,2,3,...). Aan de hand van een uitgeprint schema kunnen de medewerkers het proefveld uitzetten, zaaien en alle voor elk object specifieke behandelingen en waarnemingen uitvoeren.

Veldwaarnemingen

Gedurende het teeltseizoen verrichten IRS-medewerkers diverse waarnemingen op de proefvelden. De aard van deze waarnemingen varieert (telling of waarderingscijfers). Ook het niveau van de waarnemingen (rij of veldje) kan verschillen.

De waarnemingen worden vaak al op het veld vastgelegd in een bestand met behulp van robuuste fieldworkers (handheldcomputers). Hierin zijn door de medewerkers van de rekenkamer van te voren van elk proefveld naast het proefveldnummer de nummers van de veldjes waarin iets moet worden waargenomen in de afgesproken

loopvolgorde voorgeprogrammeerd. De kans op vergissingen is daardoor tot een minimum beperkt.

Het gebruik van de fieldworker kent een aantal beperkingen. Zo is het bijvoorbeeld niet mogelijk tijdens het doen van waarnemingen de volgorde van de veldjes aan te passen aan de looprichting van de proefveldmedewerkers (die niet altijd van te voren is in te schatten). In dat geval moeten bij het ingeven van waarnemingen in de fieldworker steeds sprongen worden gemaakt. Daarnaast vergt het vastleggen van de proefveldschema's en het uitlezen naar de PC nogal wat organisatie. Doordat het geheugen van de fieldworkers beperkt is moeten deze handelingen regelmatig plaatsvinden.

Ook bij de oogst gebruiken de proefveldmedewerkers fieldworkers. In dit geval als intermediair tussen de weegcomputer, die verbonden is met de bunker op de bietenrooier, en de PC. De weegcomputer registreert tijdens het rooien de gewichten van de afzonderlijke veldjes en de proefveldmedewerkers geven hierop het veldjesnummer in. Regelmatig worden de fieldworkergegevens uitgelezen op de PC als ASCII-bestand. Deze bestanden worden op hun beurt ingelezen in VWU, dat de waarnemingen via het proefveld- en veldjesnummer direct weet te koppelen aan de bijbehorende objecten en herhalingen.

Bietenkwaliteit

Tijdens het rooien worden bietenmonsters genomen voor verdere analyse. Belangrijk is de hoeveelheid tarra aan de bieten en het suikergehalte, maar ook het gehalte aan andere componenten die de suikerwinning beïnvloeden.

De bietenmonsters gaan in afzonderlijke zakken samen met een lezerkaart, die elk monster een uniek nummer geeft. Deze lezerkaarten zijn door VWU aangemaakt op basis van het schema van het proefveld en het ingegeven aantal gewenste monsters per veldje. De proefveldmedewerkers krijgen van de lezerkaarten een overzicht (de zogenaamde zakkenlijst), zodat ze tijdens het rooien kunnen controleren of de goede kaart bij het monster wordt gedaan.

Zoals dat ook gebruikelijk is in de suikerindustrie gaan de bietenmonsters vervolgens naar het tarreerlokaal om ze te wassen, koppen en wegen. Van elk monster wordt een hoeveelheid brij gewonnen voor verdere analyse in het laboratorium. De lezerkaart begeleidt elk bietenmonster gedurende het hele traject, vanaf binnenkomst tot aan de eindanalyse. Apparatuur die is gekoppeld aan de gewichtregistratie en de analyse-apparatuur leest dit nummer, koppelt het aan de analysegegevens en geeft dit door aan de PC. Aan het eind van de dag worden deze gegevens doorgestuurd en ingelezen in VWU.

Inlezen, valideren en aggregeren van data

Alle waarnemingsuitkomsten van de fieldworker, het tarreerlokaal en het laboratorium leest VWU via aparte procedures in. De fieldworkerbestanden bevatten de proefveld- en veldjesnummers als identificatie van de waarnemingsuitkomsten. De bestanden van het tarreerlokaal en laboratorium bevatten de lezerkaartnummers als identificatie.

Bij de analyses horen ook gegevens over metingen aan controlestandaarden, die nodig zijn om op het verloop van de meetapparatuur te corrigeren. In VWU is een procedure aanwezig die deze correctie automatisch uitvoert. Na de correctie en na de controle van het inlezen vindt validatie van de ingelezen records plaats, dat wil zeggen dat de uitkomsten gekoppeld worden aan de reeds in de database aanwezige gegevens.

Na validatie kan controle en eventuele wijziging van de gegevens plaatsvinden. Als de gegevens goed zijn bevonden worden ze geaccepteerd. Daarbij slaat VWU ze in het ene geval op als meetwaarde, of in het andere geval, als het om waarnemingen op rij of monsterniveau gaat, aggregaat het de waarden naar veldjesniveau. In VWU is van elke meetvariabele bekend hoe de aggregatie in zijn werk dient te gaan en welke formule daar bij hoort. Voor sommige meetvariabelen koppelt VWU verschillende waarnemingsuitkomsten en databasegegevens aan elkaar via een formule. Bijvoorbeeld bij het

berekenen van de suikeropbrengst per hectare: veldje-afmetingen, gewicht weegcomputer, netto en brutogewicht bepaald in het tarreerlokaal en suikergehalte bepaald in het laboratorium.

Statistische verwerking en presentatie

Voor de statistische verwerking en analyse van de gegevens beschikt VWU over een optie die de gegevens zodanig rangschikt en exporteert, dat het pakket 'Genstat' ze direct kan gebruiken.

Voor presentatie van de proefuitkomsten is een applicatie in Excel geschreven. Hierin zijn SQL-scripts ingebouwd, die de waarnemingsuitkomsten uit de database filtert en op een vooraf aangegeven manier weer geeft.

Efficiënt systeem

Met VWU beschikt het IRS over een efficiënt systeem voor de verwerking van de vele proefveldgegevens. De centrale verwerking heeft als grote voordeel dat onderzoekers weinig tijd hieraan hoeven te besteden en dit werk aan specialisten kunnen overlaten. Het systeem is wel flexibel en toegankelijk, doordat het draait op het PC-netwerk van het instituut. Onderzoekers kunnen vanuit hun werkplek toegang krijgen tot de database en hieruit data voor eigen gebruik halen.

De gegevens blijven bewaard en zullen ook in de toekomst goed toegankelijk zijn. Dit is mogelijk door de koppeling van projectnummers en een goede omschrijving van de proefomstandigheden aan de waarnemingsuitkomsten.

Voor de toekomst blijven er een aantal wensen voor verbetering. Het aantal mogelijke rapporten vanuit VWU is nu nog beperkt en vraagt om uitbreiding. Ook een integratie tussen de nu losstaande statistiek- en presentatie-programma's en VWU staat op de wensenlijst. De opzet van de database is zodanig gekozen dat dit soort verbeteringen en aanvullingen in de toekomst relatief eenvoudig te realiseren zijn. @