

Instituut voor Cultuurtechniek en Waterhuishouding
Wageningen

ASPECTEN VAN INFORMATIEVERWERKING

3

**BIBLIOTHEEK
STARINGGEBOUW**

Registratieverwerking voor automatische pF-bepalingen

ing. J.B.H.M. van Gils

Nota's van het Instituut zijn in principe interne communicatiemiddelen, dus geen officiële publikaties.

Hun inhoud varieert sterk en kan zowel betrekking hebben op een eenvoudige weergave van cijferreeksen, als op een concluderende discussie van onderzoeksresultaten. In de meeste gevallen zullen de conclusies echter van voorlopige aard zijn omdat het onderzoek nog niet is afgesloten.

Bepaalde nota's komen niet voor verspreiding buiten het Instituut in aanmerking

JSN 231292-02

A S P E C T E N V A N I N F O R M A T I E V E R W E R K I N G

Deel	Titel	Auteur	Nota	Datum
1	Computerverwerking van lange reeksen getallen	J.B.H.M. van Gils	935	nov. 1976
2	Optimaliseren van parameters: Het gereedmaken van een functie voor toepassing in NLV	Ph.Th. Stol	943	febr. 1977
3	Registratieverwerking voor automatische pF-bepalingen	J.B.H.M. van Gils	949	febr. 1977
4	Het systematisch bepalen van de afgeleiden van een functie ten behoeve van hun programmering	Ph.Th. Stol	948	febr. 1977
5	Het samenstellen van een input tape voor een elektrisch model	J.B.H.M. van Gils		
6	Over het samenstellen van een computerprogramma voor het optimaliseren van parameters	Ph.Th. Stol	951	apr. 1977

De nota's handelende over Aspecten van Informatieverwerking bevatten inlichtingen over de ontwikkeling van de informatieverwerking binnen het Instituut. Naast meer concluderende en toelichtende beschouwingen zal aandacht worden besteed aan het gebruik van programma's en programmapakketten en zullen zakelijke inlichtingen over praktijkervaring met en toepassing van de informatieverwerking worden gegeven

I N H O U D

	Blz.
INLEIDING	1
DE REGISTRATIE	2
DE FOUTEN	3
HET VERWERKINGSSYSTEEM	5
HET STANDAARDSYSTEEM	8
SLOTOPMERKINGEN	8
LITERATUUR	9
VOORBEELDEN VAN CASSETTEREGISTRATIE	

INLEIDING

In het laboratorium is een tafel ingericht met een geautomatiseerde opstelling voor de bepaling van de pF-curve en het capillair geleidingsvermogen aan hetzelfde monster. In de continu werkende opstelling kunnen maximaal 10 monsters tegelijk worden gemeten. Ieder monster staat geplaatst op een balans en kan maximaal tien tensiometers bevatten. Met een ingesteld tijdsinterval worden aan alle aangesloten monsters de gewichten en de vochtspanningen gemeten. Tevens wordt de vochtspanning van referentiewaarden gemeten. Alle metingen worden automatisch geregistreerd met behulp van een data-logger in millivolts op magnetic tape in een cassette. De geregistreeerde informatie op cassettetape kan door de PDP-11 van IWIS-TNO worden gelezen en verwerkt. De verwerking zal geschieden door met instructies gestuurde programma's volgens een daartoe opgesteld systeem. In deze nota staan de registratie en de principes van dat deel van het verwerkingssysteem beschreven dat volgens een routine procedure verloopt. Dit deel kan worden ontwikkeld uit een standaardsysteem (VAN GILS, 1977) waarvan grote delen gereed zijn. De geregistreeerde metingen kunnen hiermee in een vorm worden gebracht waardoor ze als testmateriaal gebruikt kunnen worden voor de ontwikkeling van het volgende stadium van het verwerkingssysteem.

DE REGISTRATIE

Een data-logger is een registratie-apparaat dat metingen registreert op cassettetape. Vele meetapparaten kunnen met een vast tijdsinterval worden afgelezen. De metingen worden dan achtereenvolgens op tape geregistreerd. Dan ontstaat een lange reeks karakters, een file karakters. Iedere file wordt afgesloten met een end-of-file teken. Op een cassettetape kunnen \pm 30 000 of 50 000 karakters worden geschreven, dat zijn acht bladzijden met getallen zoals voorbeeld 1 op bladzijde 10. De maximale capaciteit van de data-logger voor de pF-bepaling is één cassette in de vier dagen.

Binnen de file bestaat een indeling in records. Iedere record is afgesloten met een end-of-record teken. In de voorbeelden is iedere afgedrukte regel een record, alleen het end-of-record teken is weggelaten. De indeling van ieder record van deze logger is steeds dezelfde:

xxx	xxx	xx	xx	x	maximaal 10 keer	(xx	xxxx)
proef- nummer	dag- nummer	uur- nummer	minuut- nummer	stand v.d. vloeistof- schakelaar		kanaal- nummer	meting in millivolts

waarin xxx het aantal gereserveerde posities aangeeft. In werkelijkheid sluiten alle gereserveerde posities aan elkaar aan. Zie voorbeelden achterin. Het totaal aantal cijfers per record is dus maximaal 71.

Op de laboratoriumtafel staan maximaal 70 meters opgesteld. Een meting kan een vochtspanning van een tensiometer of een referentiemeter zijn of een gewicht van een monster. Iedere meting wordt geïdentificeerd door de klokregistratie en de stand van de vloeistofschakelaar behorende bij het record en het kanaalnummer geregistreerd bij iedere meting. Een meetronde, een scan, bestaat uit 12 records. Iedere meting komt eenmaal voor in een scan. De plaats van een meting binnen het record staat niet geheel vast. Als een kanaalnummer uitgeschakeld is, schuiven de volgende getallen naar links en wordt het record korter dan 71 karakters (zie voorbeeld 2 de records 49 t/m 54). In dit record ontbreekt kanaal 08.

Iedere meting geldt slechts voor één monster. Alleen referentie-

metingen kunnen voor meer monsters gelden. Een gewichtsmeting kleiner dan 200 millivolt betekent, dat op de betreffende drukdoos op dat moment geen monster staat. We nemen dan aan dat in dat geval een monster verwisseld wordt behalve aan het begin van de cassettetape. Een aaneensluitende reeks gewichten kleiner dan 200 millivolt wordt geacht slechts één wisseling te zijn.

DE FOUTEN

Zoals in iedere lange reeks getallen komen in deze files fouten voor. De drie registratievoorbeelden hebben erg veel fouten.

De fouten zijn het gevolg van kinderziekten van deze logger. Toch wordt bij zelfregistrerende apparaten een klein aantal fouten getolereerd. Het programmapakket dat het verwerkingssysteem bevat moet de fouten onderkennen en aangeven. De fouten mogen niet doorwerken dat wil zeggen dat moet worden voorkomen dat verdere verwerking met aanwezige fouten nieuwe fouten in het materiaal aanbrengt. Het opsporen en het verbeteren van fouten moet in een zo vroeg mogelijk stadium van de verwerking gebeuren.

De voorkomende foutsoorten kunnen worden onderkend door te toetsen op de eigenschappen van de foutsoorten:

1. Een onjuist aantal records per scan. Iedere scan behoort 12 records te hebben met de stand van de vloeistofschakelaar oplopend van 1 tot 6 en nogmaals van 1 tot 6.
2. Een onjuist aantal karakters in een record. Doordat de records een gelijke vorm hebben kan de lengte van het record worden getoetst. Zolang de recordlengte legaal verschilt, dus als kanalen in- en uitgeschakeld worden, kan niet op deze foutsoort worden getoetst.
3. Een verschuiving van karakters in het record. Deze fout wordt opgespoord als een of andere combinatie van andere foutsoorten. Niet verbeterde verschuivingsfouten werken meestal desastreus door omdat vele onjuiste getallen gevormd worden.
4. Onleesbare karakters.
5. Een onleesbare volgorde van karakters, bijvoorbeeld 92-57. Deze

foutsoort komt in de voorbeelden niet voor.

6. Een onleesbare getalsgrootte voor de PDP-11. Deze fout kan bij deze registratie niet voorkomen. Bij andere registraties komt deze fout wel voor. In dit geval duidt deze foutsoort op een verwerkingsfout.
7. Een niet-juiste getalsgrootte volgens de onderzoeker, bijvoorbeeld kanaalnummer 20.

Er bestaat een rangorde in de foutsoorten. Zolang de foutsoorten 2 en 3 niet verbeterd zijn, zullen een groot aantal foutmeldingen 5, 6 en 7 voorkomen. Wat de verbeteringen moeten zijn kan niet met het verwerkingssysteem worden bepaald. Het is de onderzoeker die de verbeteringen moet opgeven.

In voorbeeld 1 werden een drietal foutsoorten onderkend:

- (3): Verschuiving van karakters in het record komt voor in de eerste 13 records.
- (4): Een onleesbaar karakter komt voor als eerste karakter van het zevende record.
- (7): Niet juiste getalsgrootten komen voor in de eerste 13 records
namelijk de kanaalnummers 20, 30, 40, 50, 60, 70, 80, 90
het dagnummer 901
de uurnummers 46, 30, 32, 34, 40, 42, 44
de minuutnummers 65, 66, 61, 64, 65, 66, 61, 62, 63,
64, 65

De bijgeschreven omliggende cijfers zijn opgegeven verbeteringen die moeten worden voorgevoegd of tussengevoegd of die geregis- treerde karakters moeten vervangen.

T o e l i c h t i n g o p d e f o u t e n i n v o o r b e e l d 1

Voor een overzichtelijker indeling zie voorbeeld 1a. Voor de verbeterde registratie zie voorbeeld 1b.

- In het eerste record in 1a moeten alle cijfers één plaats naar rechts worden opgeschoven. Zie proefnummer dat 002 moet zijn. Zie kanaalnummers die 11, 12, ... moeten zijn. De metingen zijn een factor 10 te groot.

- In het tweede en derde record moeten eveneens alle cijfers 1 plaats worden opgeschoven.
- In het vierde record moeten alle cijfers 2 plaatsen worden opgeschoven. Zie kanaalnummers die moeten zijn 01, 02 enz.

In voorbeeld 2 worden een tweetal foutsoorten getoond:

- (1): De meeste scans hebben een onjuist aantal records.
- (2): Het aantal karakters van het 21e record is niet goed.

In voorbeeld 3 tenslotte wordt de hoeveelheid getoond waarin fouten eventueel kunnen voorkomen.

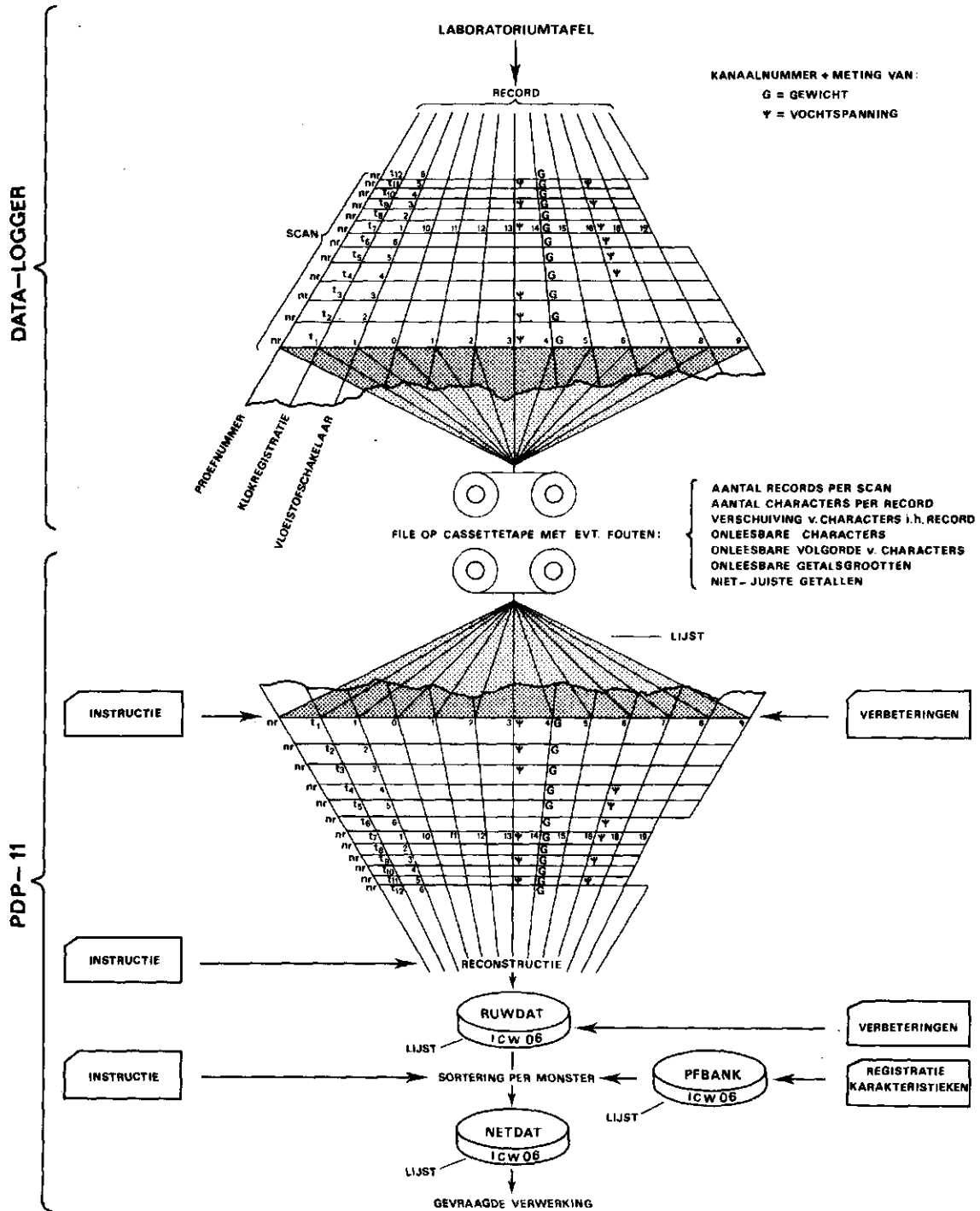
Uit de gegeven voorbeelden blijkt wel dat het opsporen en verbeteren van fouten een flinke werkbelasting geeft. Het voorkomen van fouten is het enige goede begin. Het halve werk is dan gebeurd.

HET VERWERKINGSSYSTEEM

De regelmatig terugkerende verwerking van files van zelfregistre- rende apparaten moet met een routine procedure kunnen gebeuren. Hier- voor wordt telkens een verwerkingssysteem opgesteld. De hulp van de PDP-11 geeft dan een arbeidsbesparing. De computerprogramma's zijn de gereedschappen waarmee volgens het systeem kan worden gewerkt. Fig. 1 geeft een schematisch overzicht hoe de metingen op cassettetape ver- zameld worden door de data-logger. Iedere horizontale lijn stelt een record voor. Een serie van 12 records vormt een scan waarin alle me- tingen voorkomen. De eerste 6 records bevatten de metingen van kanaal 0 tot en met 9 met hun identificaties. De volgende 6 records bevat- ten op dezelfde wijze de metingen van kanaal 10 tot en met 19. De identificatie van de metingen wordt aan de laboratoriumtafel geno- teerd en per monster op ponskaarten verzameld met alle andere regis- tratiekaracteristieken.

De computerbewerkingen worden uitgevoerd door met instructies ge- stuurde programma's. Ik onderscheid de verwerking in vier gedeelten:

Fig. 1. Weergave van de registratieverwerking van één monster bij pF-bepaling



1. De verwerking van de karakters. De programma's lezen de files met karakters van de cassettape, schrijven deze files op schijf, maken lijsten, sporen de foutsoorten 2 en 3 op en schrijven de verbeteringen in de files. Doordat de records een gelijke vorm hebben kan ieder record apart verwerkt worden. Iedere meting is geïdentificeerd binnen het record. Het optreden van foutsoort 1 resulteert alleen in een aantal onbekende metingen. Foutsoort 1 moet wel opgespoord worden omdat meetfouten of registratiefouten optreden en verbeteringen van de gemeten gegevens moeten worden opgegeven. Dit gebeurt door de onbekende metingen op te sporen.
2. De reconstructie van het registratieschema. Uit de karakters worden nu getallen gevormd. Daarbij worden de foutsoorten 4, 5, 6 en 7 opgespoord. De getallen worden geschreven op file 'RUWDAT' gegroepeerd volgens het schema waaruit de cassettape door de data-logger werd beschreven. Vervolgens wordt een lijst gemaakt en worden de verbeteringen in file 'RUWDAT' aangebracht.
3. De sortering van de getallen per monster. De registratiekarakteristieken worden per monster opgeborgen in file 'PFBANK'. Van file 'PFBANK' worden regelmatig tabellen gemaakt. De opgegeven identificatie van de metingen wordt gebruikt om de getallen per monster te verzamelen uit file 'RUWDAT' en gesorteerd per monster te schrijven in file 'NETDAT'. Desgewenst kan foutsoort 7 nogmaals worden opgespoord per monster en worden verbeterd.
4. De door de onderzoeker gevraagde verwerking. De beschrijving van deze verwerking is nog niet in het stadium dat hiervoor alle programma's kunnen worden gemaakt. In ieder geval zullen de metingen gemeten in millivolts moeten worden getransformeerd tot voor de gebruiker bekende fysische grootheden. De uiteindelijk bereikte informatie wordt per monster opgeborgen in file 'PFBANK'. De bedoeling is dat deze file een 'databank' voor pF-gegevens wordt.

HET STANDAARDSYSTEEM

De indeling van de files van deze logger is complex. De meeste andere mij bekende reeksen getallen hebben een eenvoudiger indeling. Het systeem is dan ook geldig voor de verwerking van vele soorten reeksen getallen zoals enquêtes, peilen, boekhoudkundige gegevens en verdampingsgegevens. Iedere registratievorm heeft een eigen ingang in het systeem. Voor enkele registratievormen moet het systeem nog worden uitgebreid. Voor de sorteringsbewerking is nog geen standaard methode gevonden. Voor de specifieke verwerking volgens opgave van de onderzoeker kunnen programma's worden toegevoegd.

Het standaardsysteem is opgebouwd uit programma's die kunnen lezen en schrijven in een standaardvorm van opslag op schijf. Het bevat ook programma's die kruistabellen maken en berekeningen met lineaire functies uitvoeren. Daarnaast is een programma beschikbaar om getallen vanuit de standaardvorm te converteren naar een vorm identiek aan de ponskaartregistratie zodat iedere bewerking met ieder programma en met iedere computer tot de mogelijkheden behoort.

SLOTOPMERKINGEN

Het verwerkingssysteem is geïntroduceerd in ICW-nota 935. Bij ieder programma is een gebruiksinstructie gevoegd. Een algemene beschrijving van de programma's is nog niet gerealiseerd.

Een op routinegebruik gebaseerd verwerkingssysteem behoeft een praktische handleiding. Deze verschilt in ieder geval per registratievorm. Veelal zullen zelfregistrerende apparaten ieder een eigen uitgebreide handleiding voor de verwerking nodig hebben, omdat de verwerking gedurende meer jaren zal plaatsvinden. Hiervan is momenteel nog niets gerealiseerd.

Het verwerkingssysteem is zo ingericht dat iedere ICW-medewerker dit, eventueel met behulp van de afdeling Wiskunde, vrijelijk kan gebruiken en zelfstandig en naar eigen inzicht zijn gegevenbestanden kan beheren.

LITERATUUR

GILS, J.B.H.M. VAN, 1977. Computerverwerking van lange reeksen getal-
len. ICW-nota 935.

Voorbeeld 1a. De registratie met fouten van voorbeeld 1

Record	Proef	Dag	Uur	Minuut	Schake- jaar	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting						
1	021	901	04	65	1	00	7501	10	4321	20	0000	30	0651	40	0005	00	0060	00	0700	00	1800	00	0900	61	
2	021	901	10	66	1	00	6411	10	4331	20	0171	30	0651	40	0311	50	0001	60	0001	70	0001	80	0001	90	061
3	21	901	12	61	0	00	7570	10	4270	20	0190	30	0590	40	0340	50	4290	60	0000	70	0000	80	0000	90	067
4	319	011	46	20	0	06	5001	04	2702	00	2303	00	5904	00	3505	04	2906	00	0007	00	0008	00	0009	00	67
5	219	012	06	30	0	05	7301	04	2702	00	2603	00	5904	00	3705	04	2906	00	0007	00	0008	00	0009	00	67
6	021	901	22	64	0	00	0500	10	4270	20	0350	30	0590	40	0530	50	4290	60	0000	70	0000	80	0000	90	067
7	21	901	24	65	0	00	0500	10	4270	20	0350	30	0590	40	0510	50	4290	60	0000	70	0000	80	0000	90	067
8	021	901	30	66	0	00	3610	10	4270	20	0300	30	0590	40	0420	50	4290	60	0000	70	0000	80	0000	90	067
9	021	901	32	61	1	02	0751	10	4311	20	0001	30	0651	40	0001	50	0001	60	0001	70	0001	80	0001	90	061
10	021	901	34	62	1	01	0751	10	4311	20	0011	30	0651	40	0131	50	0001	60	0001	70	0001	80	0001	90	061
11	021	901	34	62	1	01	0751	10	4311	20	0011	30	0651	40	0281	50	0001	60	0001	70	0001	80	0001	90	061
12	021	901	42	64	1	01	7231	10	4311	20	0001	30	0651	40	0001	50	0001	60	0001	70	0001	80	0001	90	061
13	021	901	44	65	1	00	8751	10	4311	20	0031	30	0651	40	0191	50	0001	60	0001	70	0001	80	0001	90	061
14	002	190	15	06	6	10	0697	11	0431	12	0015	13	0065	14	0034	15	0000	16	0000	17	0000	18	0000	19	0061
15	002	190	15	26	1	00	0842	01	0427	02	0019	03	0059	04	0030	05	0427	06	0000	07	0000	08	0000	09	0067

Voorbeeld 1b. De verbeterde registratie van voorbeeld 1

Record	Proef	Dag	Uur	Minuut	Schake- jaar	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting	Kanaal	Meting						
1	002	190	10	46	5	10	0750	11	0432	12	0000	13	0065	14	0000	15	0000	16	0000	17	0000	18	0000	19	0061
2	002	190	11	06	6	10	0641	11	0433	12	0017	13	0065	14	0031	15	0000	16	0000	17	0000	18	0000	19	0061
3	002	190	11	26	1	00	0757	01	0427	02	0019	03	0059	04	0034	05	0429	06	0000	07	0000	08	0000	09	0067
4	002	190	11	46	2	00	0650	01	0427	02	0023	03	0059	04	0035	05	0429	06	0000	07	0000	08	0000	09	0067
5	002	190	12	06	3	00	0573	01	0427	02	0026	03	0059	04	0037	05	0429	06	0000	07	0000	08	0000	09	0067
6	002	190	12	26	4	00	0050	01	0427	02	0035	03	0059	04	0053	05	0429	06	0000	07	0000	08	0000	09	0067
7	002	190	12	46	5	00	0050	01	0427	02	0035	03	0059	04	0051	05	0429	06	0000	07	0000	08	0000	09	0067
8	002	190	13	06	6	00	0361	01	0427	02	0030	03	0059	04	0042	05	0429	06	0000	07	0000	08	0000	09	0067
9	002	190	13	26	1	10	2075	11	0431	12	0000	13	0065	14	0000	15	0000	16	0000	17	0000	18	0000	19	0061
10	002	190	13	46	2	10	1075	11	0431	12	0001	13	0065	14	0013	15	0000	16	0000	17	0000	18	0000	19	0061
11	002	190	14	06	3	10	0914	11	0431	12	0011	13	0065	14	0028	15	0000	16	0000	17	0000	18	0000	19	0061
12	002	190	14	26	4	10	1723	11	0431	12	0000	13	0065	14	0000	15	0000	16	0000	17	0000	18	0000	19	0061
13	002	190	14	46	5	10	0875	11	0431	12	0003	13	0065	14	0019	15	0000	16	0000	17	0000	18	0000	19	0061
14	002	190	15	06	6	10	0697	11	0431	12	0015	13	0065	14	0034	15	0000	16	0000	17	0000	18	0000	19	0061
15	002	190	15	26	1	00	0842	01	0427	02	0019	03	0059	04	0030	05	0427	06	0000	07	0000	08	0000	09	0067

