

RAPORT ȘTIINȚIFIC

Contract nr 33CI/2017, cod PN-III-P2-2.1-CI-2017-0654

Titlu proiect: *Sistem integrat de analiză și prognoză a consumului pentru distribuitorii IMM de gaze naturale (FOReGASt)*

Durata proiect: 5luni;

I. Activități de cercetare experimentală

Activitățile de cercetare întreprinse au avut ca scop dezvoltarea de modele predictive pe două niveluri:

1. Modele care prezic consumul lunar la nivelul fiecărui consumator
2. Modele care prezic consumul total lunar la nivelul fiecărei localități și a fiecărei categorii de consumatori.

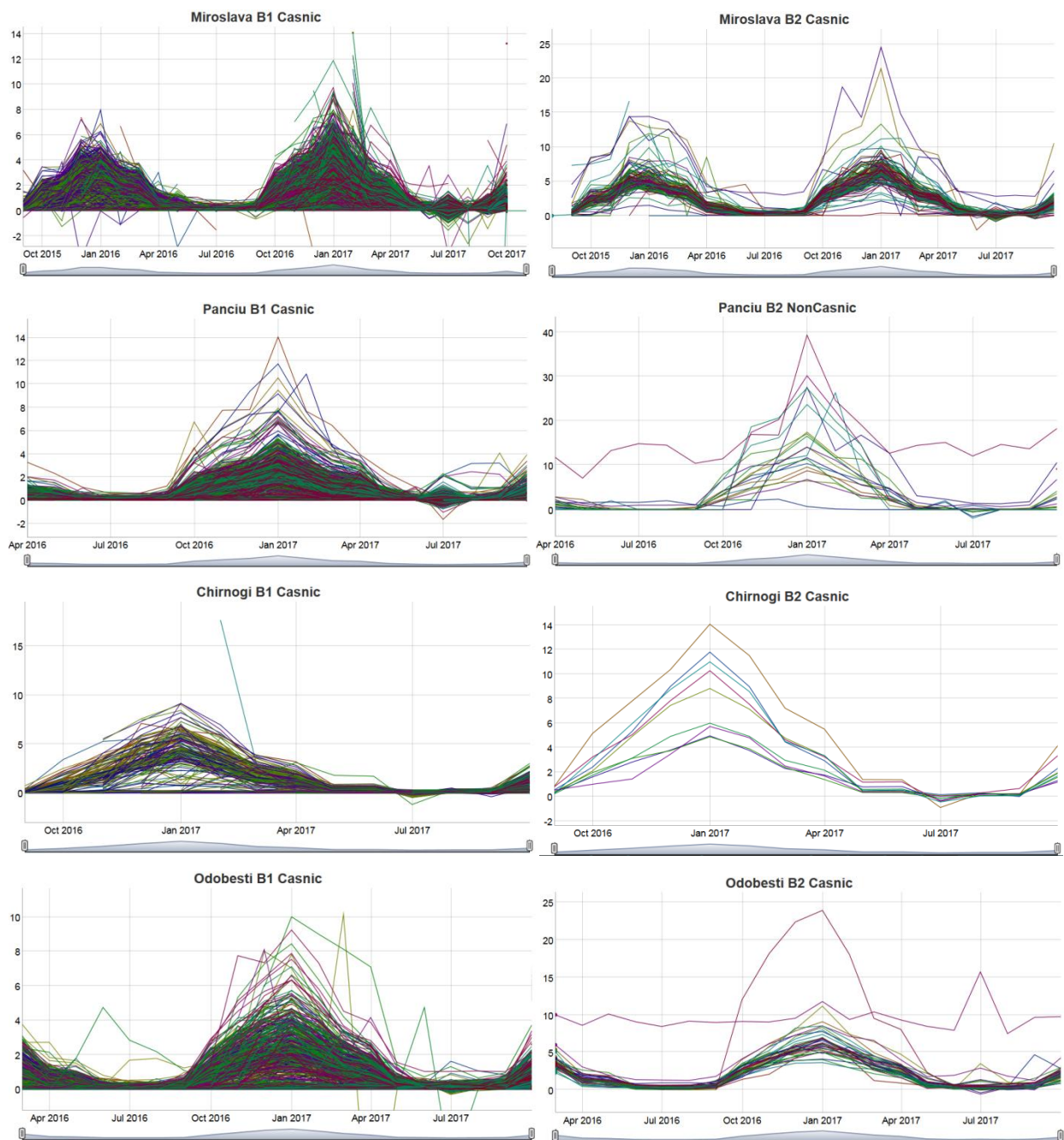
Ambele direcții au apelat la utilizarea metodelor statistice de tip ARIMA si ARIMA hibride (cu variabile exogene) și a unui spectru mai larg de metode de învățare automată.

Datele experimentale au constat din consumurile lunare ale clienților din 4 localități. Distribuția acestora pe categorii de consum este ilustrată în continuare:

	Miroslava	Panciu	Chirnogi	Odobesti	Total
B1-NonCasnic	117	46	14	27	204
B1-Casnic	1690	911	267	631	3499
B2-NonCasnic	19	20	0	12	51
B2-Casnic	121	37	9	38	205
B3-NonCasnic	6	15	0	5	26
B3-Casnic	0	1	0	6	7
B4-NonCasnic	1	0	0	2	3
B4-Casnic	0	0	0	1	1
Total	1954	1030	290	722	3996

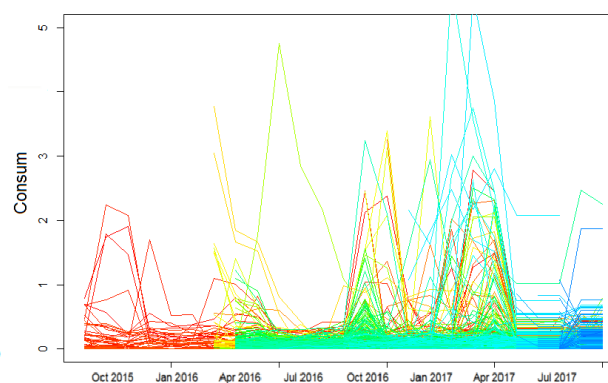
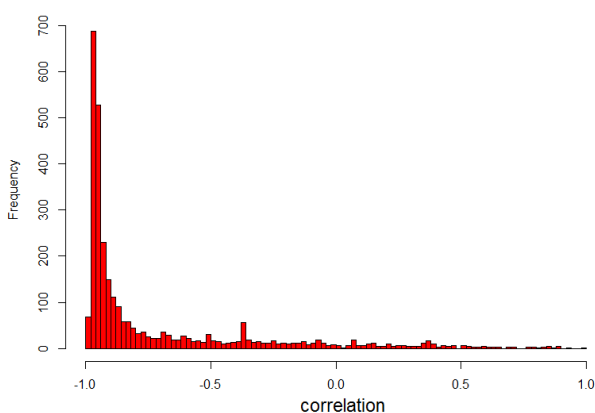
Istoricul diferă ca durată pentru cele patru localități, cele mai vechi date fiind cele corespunzătoare localității Miroslava cu înregistrări începând din septembrie 2015. Următoarele

grafice ilustrează ca serii de timp consumurile în MWh pentru cele mai bine reprezentate clase de consumatori.



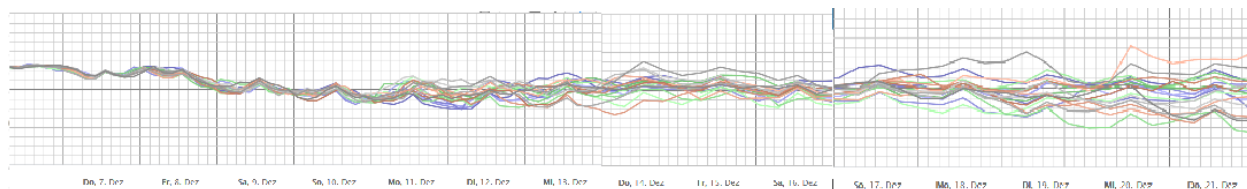
După cum relevă și graficele de consum, datele au necesitat o prelucrare prealabilă pentru a elimina/corecta consumurile negative rezultat a unor înregistrări eronate. De asemenea, în anumite luni operațiunea de citire a contorului a fost înlocuită de estimări realizate de operator, caz în care erorile se corectează odată cu citirile ulterioare.

Date meteorologice istorice pentru localitățile în cauză au fost extrase din raportările ANM ce cuprind valori zilnice ale temperaturii (accesibile la adresa <https://www.ogimet.com/gsocd.phtml.en>). Stațiile de referință au fost identificate apelând la specialiști în meteorologie din cadrul Facultății de Geografie și Geologie a Universității "Alexandru Ioan Cuza" din Iași, pentru a surprinde particularitățile climatice a zonelor în analiză (în acest sens nu doar distanța între punctul de interes și stații joacă un rol important ci și formele de relief comune sau care se interpun între stație și localitatea de interes). Datele de la următoarele stații au fost preluate pentru a fi incluse în analiză, după cum urmează: Miroslava - stația Iași, Chirnogi – stația Oltenița; Panciu și Odobesti – stația Focșani. Pe baza temperaturilor zilnice raportate au fost extrase statistici la nivel de lună: minime, maxime, mediane și medii. Temperatura este principalul factor extern care influențează consumul: pentru categoria consumatorilor casnici de tip B1, corelația Pearson calculată între consumul mediu lunar și temperatura medie lunară pentru cele 25 de luni înregistrate este de -0.96, în timp ce pentru celelalte categorii de consumatori casnici depășește în modul -0.97. Corelația liniară între consumul mediu și temperatura medie lunară este semnificativă și pentru diferitele categorii de consumatori non-casnici (deși mai redusă în modul), cu excepția singurului consumator existent de tip B4 pentru care consumul nu pare a fi influențat de condițiile meteorologice. Corelația între consumul lunar al unui consumator și temperatura medie lunară constituie unul dintre criteriile suplimentare de împărțire a clienților în categorii pentru care sunt derivate modele diferite. Histograma de mai jos stânga ilustrează distribuția valorii corelației liniare între consumul lunar și temperatura medie lunară pentru toți consumatorii din baza de date, în timp ce graficul din dreapta ilustrează consumurile lunare înregistrate pentru consumatorii cu corelații negative ne semnificative și pentru cei cu corelații pozitive.



Pe baza corelației între consum și temperatura, clienții au fost împărțiți în două categorii: cei cu corelații negative semnificative (însușează un total de 2396 clienți casnici și 153 non-casnici – la un nivel de semnificație de 0.1) și cei cu corelații nesemnificative și semnificativ pozitive (aceștia corespund la clienți ce nu utilizează gazul pentru încălzire sau clienți care nu apelează continuu la această sursă de energie – de ex. casele de vacanță).

Predicții cu acuratețe crescută a consumului pentru momente viitoare depind de prognoze realiste ale temperaturii. Prognozele meteo pierd însă din acuratețe odată cu distanța în timp, motiv pentru care se consideră a fi de încredere prognoze pe cel mult 10 zile calendaristice. Prognozele meteo se realizează utilizând o serie de modele matematice; graficul de mai jos ilustrează cum acestea diverg odată cu deplasarea pe axa timpului în raport cu momentul de timp curent (sursa: <https://kachelmannwetter.com>).



Datorită acestor limitări, modele de predicție a consumului au fost dezvoltate și evaluate în trei situații diferite: cazul ideal în care se dispune de temperaturile reale la nivelul lunii pentru care se prezice, cazul în care se dispune de prognoza pentru primele 10 zile ale lunii și cazul în care nu se dispune de aceste date externe iar predicția este bazată doar pe valorile anterioare ale consumului – situație specifică analizei seriilor de timp. Prognoze meteo pot fi obținute automat contra-cost sau gratuit cu ajutorul unor API puse la dispoziție de o serie de furnizori: OpenWeatherMap sau AccuWeather furnizează gratuit prognoze pentru 5 zile și contra-cost prognoze pentru 16 zile, TheWeatherChannel furnizează gratuit prognoza pe 3 zile și contra-cost pe 10 zile iar APIXU Weather API furnizează gratuit prognoza pentru 10 zile cu o limită a numărului de interogări lunare de până la 5000. Existența unui API pentru a extrage automat prognoza meteo în timp real în cadrul aplicației adresată OSD este necesară.

1. Predicția consumului lunar la nivel de consumator

Algoritmii analizați în acest context sunt:

1. Regresia liniară standard și quantilică
2. Rețelele neuronale
3. Regresia cu vectori suport

4. Modele de tip autoregresiv cu medie mobilă (ARIMA)
5. Modele de tip autoregresiv cu variabile exogene (ARMAX)

Variabilele cu rol de predictor introduse în analiză sunt observațiile înregistrate la momente anterioare de timp, agregări ale acestor date istorice și date despre temperatură. Determinarea lungimii istoricului inclus în analiză s-a făcut empiric. S-au antrenat modele independente pentru fiecare categorie de consum, dar și modele comune în care categoria de consum a fost utilizată ca variabilă cu rol de predictor.

Deși corelația liniară între valorile consumului și temperaturile lunare sunt semnificative, modelele neliniare sunt justificate și de studii anterioare raportate pe piața energetică românească care identifică funcția sigmoidă ca fiind cea mai reprezentativă pentru a defini legătura dintre consum și temperatură [Mihai Albușescu, Sorin Neacsu. “Principii moderne privind alocările de gaze în vederea echilibrării rețelei” – proiect dezvoltat în parteneriat E.ON și Universitatea Petrol și Gaze Ploiești].

Pentru evaluarea modelelor s-au utilizat următoarele metrici: eroarea pătratică medie, eroarea absolută medie, eroarea absolută mediană, corelația liniară între valorile prezise de model și valorile reale.

Performanța modelelor 1-3 variază de la o lună la alta, iar durata istoricului utilizat în predicție a trebuit adaptat în funcție de luna pentru care se realizează prognoza. În plus, datorită faptului că noi consumatori se adaugă în sistem, și antrenarea de modele ce utilizează număr redus de valori istorice de consum au fost necesare. Experimentele au confirmat ipoteza conform căreia includerea temperaturilor reale ca predictor crește semnificativ acuratețea predicției – însă doar pentru predicțiile realizate pentru lunile septembrie – aprilie. În cazul utilizării prognozei meteo pe primele 10 zile calendaristice ale lunii, eroarea globală calculată cumulat pe toate lunile este redusă comparativ cu modelele ce nu utilizează deloc date de temperatură, însă există situații punctuale în care eroarea lunară este mai mare.

2. Predicții ale consumului lunar total

Metodologia pentru a prezice consumul total a urmat trei direcții:

1. Estimarea directă a consumului total incluzând ca variabile cu rol de predictor numărul de consumatori, istoricul de consum, condițiile de temperatură; abordări similare se găsesc raportate în literatură pentru estimarea consumului total zilnic [DEMIREL,

ÖMER FAHRETTİN, et al. "Forecasting natural gas consumption in Istanbul using neural networks and multivariate time series methods." Turkish Journal of Electrical Engineering & Computer Sciences 20.5 (2012): 695-711.];

2. Estimarea consumului mediu lunar pe categorii de consum și scalarea acestuia la numărul de consumatori; variabile cu rol de predictor utilizate sunt istoricul de consum și condițiile de temperatură;
3. Calculul consumului total bazat pe estimările realizate individual la nivel de client în cadrul primei direcții de cercetare.

II. Soluția software

Aplicația software integrează mai multe modele de predicție.

La nivelul clientului, utilizatorul are posibilitatea de a vizualiza istoricul de consum dar și predicțiile pentru perioada viitoare obținute cu un ansamblu de algoritmi.

La nivelul unei localități vor putea fi vizualizate datele istorice pe categorii de consumatori precum și prognoze realizate de un ansamblu bazat pe metodele prezentate mai sus.

O parte dintre algoritmi de predicție necesită reantrenare periodică pentru a se adapta la noile profiluri de consum, sarcină ce îi va reveni beneficiarului cecului de inovare. Furnizorul de servicii va oferi asistență în acest scop.

Soluția software dezvoltată este o aplicație web realizată în Shiny RStudio, codul fiind scris integral în R. Aplicația beneficiarului în care va fi integrată este o aplicație web gazduită în cloud, integrarea noii componente fiind astfel ușor de realizat. Este necesară conectarea la serverul de baze de date și integrarea cu un modul de securitate care nu este livrat de către furnizor.

III. Valorificarea rezultatelor cecului de inovare

Activitățile desfășurate în cadrul cecului de inovare au implicat un volum de muncă consistent investit în cercetare experimentală, pentru a integra o serie de algoritmi de învățare automată într-un ansamblu cu acuratețe sporită. Echipa furnizorului de servicii va monitoriza performanța

sistemului de predicție obținut în lunile ce urmează. Metodologia utilizată și rezultatele obținute pot face obiectul unor brevete și/sau publicații științifice.

Beneficiarul, cu asistența furnizorului de servicii, va evalua sistemul în mod constant și va estima împreună cu operatorul de sistem de distribuție eficacitatea sa în contextul politicii contractuale întreprinse de acesta. Beneficiarul va putea promova ulterior produsul software către alți potențiali clienți OSD.

Actualul cec de inovare a pus bazele unei colaborări durabile între compania beneficiar și unitatea de cercetare furnizor de servicii, deschizând drumul pentru cercetări viitoare comune ce vizează integrarea contoarelor inteligente cu predicții zilnice și recomandări adresate consumatorilor.

IV. Gradul de realizare a rezultatelor estimate în ofertă

Obiectivele estimate au fost atinse, livrabilele prevăzute fiind realizate.

Acestea au constat în principal într-o suită de algoritmi care să ofere predicții de acuratețe ridicată a consumului la nivel de client și de grupuri de consumatori. Un pachet software a fost realizat pentru a asista OSD în activitatea de nominalizare a consumului și rezervare de capacitate.

V. Indicatorii de stare și progres

prevăzuți sunt derivați direct din segmentul operațional pe care îl adresează aplicația de prognoză dezvoltată.

Durata operațiunilor de estimare a capacității și consumului după integrarea modulului de predicție este mult diminuată, pachetul software oferind facilități de vizualizare a datelor istorice și estimări ale consumului viitor.

Gradul de reducere a dezechilibrelor între volumul nominalizat și cel realizat este semnificativ; de exemplu, pentru luna octombrie 2017 abaterea absolută realizată de modelul de predicție a consumului total pentru categoria clienților casnici a constituit 51% din abaterea înregistrată de algoritmul naiv ce face estimări bazat pe aceeași lună a anului anterior.

Acuratețea crescută a prognozelor diminuează costurile OSD; acestea urmează a fi monitorizate pe parcursul mai multor luni după integrarea modulului de predicție în aplicația beneficiarului.

8 Decembrie 2017

Contractor - Furnizor de servicii

Rector/Director general

Ordonator de credite

Onofrei Mihaela

Director economic

Iftimia Liliana

Responsabil Proiect

Breabăn Mihaela Elena

Avizat,

Contractor - Beneficiar

Director general

Sârbu Alexandru-Gabriel

Director economic

Munteanu Natalia

Responsabil Proiect

Sârbu Alexandru Gabriel

Program 2: Creșterea competitivității economiei românești prin cercetare, dezvoltare și inovare
 Subprogramul 2.1. Competitivitate prin cercetare, dezvoltare și inovare – Cecuri de inovare

Indicator de rezultat	UM procent/ numar	Cantitate
Beneficiar (întreprindere mică sau mijlocie)	Nr.	1
Valoarea contribuției financiare la proiect a beneficiarului	Nr.	5000
Ponderea contribuției financiare la proiect a beneficiarului	%	10%
Brevete solicitate la nivel național și internațional	Nr.	0
Cereri de înregistrare la ORDA	Nr.	0
Produse/Tehnologii/Metode/Servicii noi sau semnificativ îmbunătățite, realizate (număr, pe categorii)	Nr.	-Produse software semnificativ îmbunătățite: 1 -Metode algoritmice integrate, dedicate problemei investigate: 1