

## **Predicció del nivell d'ocupació de l'àrea DUM a la ciutat de Barcelona**

Fase 1. Disseny i validació del model predictiu d'ocupació

Fase 2. Construcció de l'algoritme i funcional per cridar  
les dades i estimar el model predictiu a la zona pilot

## Contingut

1. Introducció.....	3
2. Definició de la zona pilot .....	4
3. Interfície de l'algoritme VBA.....	6
4. Flux de l'algoritme VBA.....	10
4.1. Entrada al sistema de les dades de registre d'usuaris DUM .....	10
4.2. Generació d'hora de sortida als registres amb hora de sortida inexistent o no vàlida.....	11
4.3. Generació de la sèrie temporal d'ocupació registrada per a cada subàmbit.....	13
4.4. Càlcul de les variables associades a la ST d'ocupació (regressors).....	15
4.5. Estimació dels models de sèrie temporal.....	17
4.6. Prediccions de l'ocupació registrada en base als models de sèrie temporal.....	18
4.7. Correcció de les prediccions per introduir l'ocupació no registrada.....	20
4.8. Càlcul del percentatge d'ocupació .....	21
4.9. Transformació del percentatge d'ocupació en semàfor verd, groc o vermell ....	22
5. Comentaris sobre línies de futur .....	22

## 1. Introducció

La recent implantació de l'Àrea DUM (Descàrrega Urbana de Mercaderies) a la ciutat de Barcelona fa necessari un sistema de predicció de l'ocupació. Aquest sistema ha de permetre als usuaris de vehicles comercials conèixer amb antel·lació si serà més o menys probable que trobin lloc per estacionar en una determinada zona DUM. BSM ha encarregat a l'Institut de Recerca en Economia Aplicada IREA-Universitat de Barcelona el disseny d'un algoritme en Visual Basic for Applications (VBA) capaç de predir aquesta ocupació a temps real.

Inicialment, aquest projecte s'aplica a una zona pilot i, en pocs mesos, s'estendrà a la resta de la ciutat. La zona pilot està formada per quatre àmbits: Plaça Catalunya, Plaça Urquinaona, Plaça Letamendi, i Calàbria. Cada àmbit s'ha dividit en 3-5 subàmbits, que agrupen entre 1 i 3 zones DUM. Els subàmbits seran la unitat d'anàlisi, en el sentit que les prediccions d'ocupació es donaran per subàmbit, no per zona DUM individual. Pel que fa a la dimensió temporal, l'anàlisi es limita de dilluns a divendres, restant els dissabtes pendents de ser incorporats.

L'objectiu d'aquest document és presentar l'algoritme per a la predicció de l'ocupació de l'Àrea DUM. Aquest algoritme incorpora com a input el flux de registres d'usuaris DUM que es va alimentant de forma contínua a mida que els usuaris utilitzen l'aplicació de registre. Tot i l'obligatorietat del registre, no tots els usuaris fan servir l'aplicació. També hi ha molts usuaris que registren l'arribada, però no la sortida de l'estacionament. Per tal de completar la informació que deriva del registre, BSM va dur a terme al 2016 diferents treballs de camp, on es van recollir dades d'ocupació real i de durada de l'estada, que han permès suplir en part les deficiències del registre d'usuaris.

Així, en base al registre d'usuaris, l'algoritme construeix una sèrie temporal amb periodicitat 5 minuts, que recull el nombre de places ocupades a cada subàmbit. Cada 5 minuts, l'algoritme calcula una nova observació d'aquesta sèrie temporal. Un cop construïda la sèrie, l'algoritme estima un model estocàstic de sèries temporals, concretament un model ARIMA (Autorregressive Integrated Mean Average). Aquest model té en compte l'ocupació dels darrers minuts, i dades sobre dia de la setmana, franja horària, etc., i és capaç de predir l'ocupació des del mateix moment en què es fa la petició, fins a 48 hores després. Quant menys temps hi hagi entre el moment en què l'usuari fa la petició i el moment en què vol estacionar, més fiable serà la predicció. Aquests models s'aniran ajustant després de la posta en producció d'aquesta primera fase.

L'algoritme finalitza amb el càlcul del percentatge estimat d'ocupació de les places operatives. Posteriorment, BSM implementarà un aplicatiu adreçat a l'usuari de l'Àrea DUM, que haurà de mostrar-li sota petició un semàfor verd, groc o vermell, segons el nivell d'ocupació previst a l'hora que l'usuari vol estacionar.

En els següents apartats es defineix la zona pilot, i es mostra detalladament el funcionament i el flux de l'algoritme de predicció. Es tracta d'un algoritme en VBA, que crida un executable del programa estadístic R.

## 2. Definició de la zona pilot

En una primer fase, l'estudi d'ocupació de l'àrea DUM es posa en marxa a l'anomenada Zona Pilot. Aquesta zona correspon a quatre àmbits: 1. Plaça Urquinaona; 2. Plaça Catalunya; 3. Plaça Letamendi; 4. Carrer Calàbria.

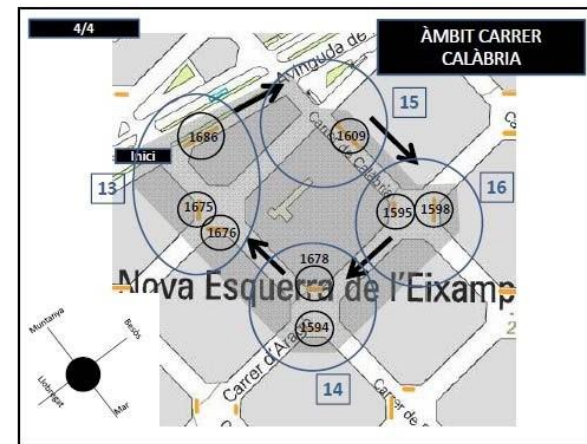
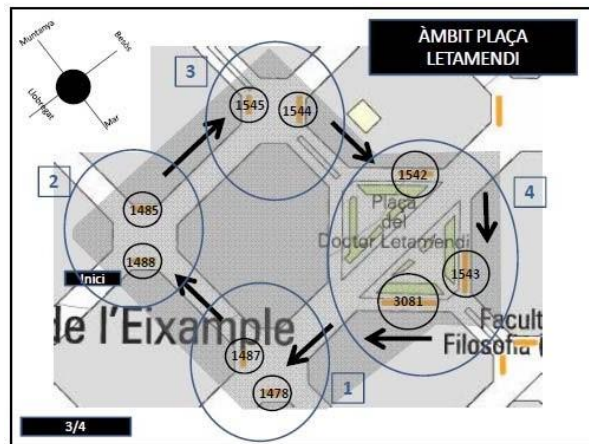
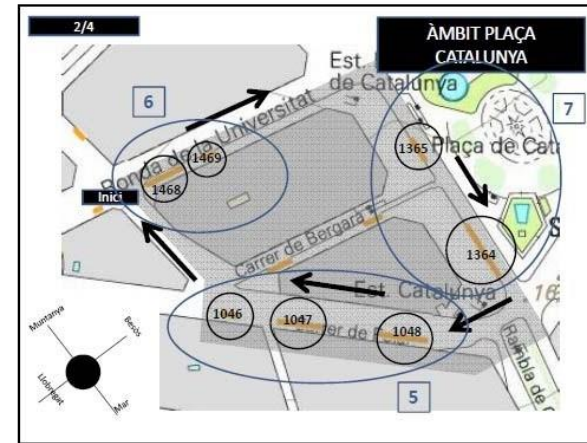
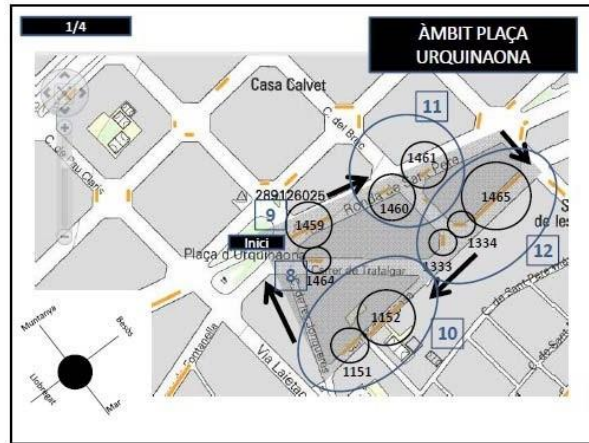
Cada àmbit es divideix en 3-5 subàmbits, formats per entre 1 i 3 zones DUM properes. A tots els efectes, **els subàmbits seran les unitats d'anàlisi d'aquest estudi**. Les sèries històriques d'ocupació i les prediccions correspondran a subàmbits, no a zones DUM individuals. Així, un usuari que desitgi estacionar a una zona DUM en particular es considerarà que fa la petició relativa al subàmbit que conté aquesta zona DUM. I la predicció de l'ocupació que se li proporcionarà correspondrà igualment a la totalitat del subàmbit.

**Taula 1.** Correspondència entre àmbits, subàmbits i zones DUM.

ID_ZONADUM	ÀMBIT	SUBÀMBIT	PLACES OPERATIVES*
1464	1	8	3
1459	1	9	11
1151	1	10	3
1152	1	10	9
1460	1	11	4
1461	1	11	8
1333	1	12	6
1334	1	12	4
1465	1	12	8
1046	2	5	2
1047	2	5	5
1048	2	5	6
1468	2	6	2
1469	2	6	11
1364	2	7	5
1365	2	7	3
1478	3	1	4
1487	3	1	5
1485	3	2	5
1488	3	2	5
1544	3	3	5
1545	3	3	5
1542	3	4	5
1543	3	4	4
3081	3	4	5
1675	4	13	5
1676	4	13	5
1686	4	13	5
1594	4	14	5
1678	4	14	4
1609	4	15	5
1595	4	16	5
1598	4	16	5

\*Places operatives a 26/11/16. Aquesta dada s'haurà d'actualitzar diàriament.

Figura 1. Zona pilot, amb àmbits i subàmbits



### 3. Interfície de l'algoritme VBA

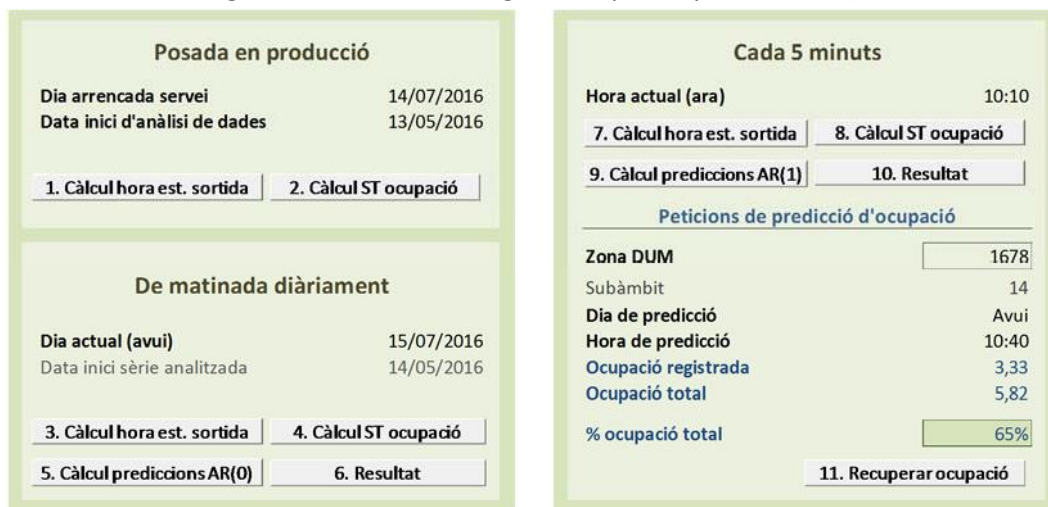
L'algoritme es troba en un full Excel, que serveix perquè BSM vegi quin és el seu funcionament. Permet fer proves amb zones DUM reals, i dies i hores simulats. Això facilitarà la comprensió de l'algoritme i la seva traducció al llenguatge que correspongui. Aquest full Excel té diverses pestanyes, que passem a descriure breument. A l'apartat sobre el flux de l'algoritme es defineixen més a fons els conceptes i processos que ara s'esbossen, de forma que recomanem llegir-lo per entendre **què es fa i per què**. L'algoritme que es presenta està dissenyat per funcionar per a tots els subàmbits. Tanmateix, per evitar esperes en els càlculs mentre es fan les proves, només s'implementen aquí els subàmbits 1 i 14.

a) Pestanya DADES\_TOT: conté 5 columnes. Les 3 primeres procedeixen del registre d'usuaris DUM. Com a mínim, aquesta pestanya ha de contenir el registre dels dos mesos anteriors a l'arrencada del servei, finalitzant el dia anterior a l'arrencada.

- FHSTART: Data i hora de registre d'estacionament de l'usuari DUM (l'usuari acaba d'estacionar), a través de SMS o de l'aplicació d'Àrea DUM.
- FHSTOP: Data i hora de registre de sortida de l'usuari DUM.
- ID\_ZONADUM: Identificador de la zona DUM, segons codis recollits a la Taula 1.
- SUBÀMBIT: Subconjunt de zones DUM a què correspon la zona identificada amb ID\_ZONADUM. A efectes d'aquest algoritme, i per no allargar de forma innecessària els temps de càlcul, només s'inclouen els subàmbits 1 i 14. Tanmateix, l'algoritme funcionarà igualment quan s'incorpori el registre de tots els subàmbits de la zona pilot.
- FHSTOP\_NOVA: Data i hora de registre de sortida de l'usuari DUM, quan la FHSTOP és incorrecta/inexistent i ha estat necessari recalculer-la. Inicialment aquesta columna està buida. S'omplirà quan s'executi l'algoritme.

b) Pestanya PRINCIPAL: es tracta pròpiament de la interfície de l'algoritme.

**Figura 2.** Interfície de l'algoritme (pestanya PRINCIPAL)



The interface is divided into two main sections:

- Posada en producció:**
  - Inputs: Dia arrencada servei (14/07/2016), Data inici d'anàlisi de dades (13/05/2016).
  - Buttons: 1. Càlcul hora est. sortida, 2. Càlcul ST ocupació.
- De matinada diàriament:**
  - Inputs: Dia actual (avui) (15/07/2016), Data inici sèrie analitzada (14/05/2016).
  - Buttons: 3. Càlcul hora est. sortida, 4. Càlcul ST ocupació, 5. Càlcul prediccions AR(0), 6. Resultat.
- Cada 5 minuts:**
  - Input: Hora actual (ara) (10:10).
  - Buttons: 7. Càlcul hora est. sortida, 8. Càlcul ST ocupació, 9. Càlcul prediccions AR(1), 10. Resultat.
  - Peticions de predicció d'ocupació:**
    - Zona DUM: 1678
    - Subàmbit: 14
    - Dia de predicció: Avui
    - Hora de predicció: 10:40
    - Ocupació registrada: 3,33
    - Ocupació total: 5,82
    - % ocupació total: 65%
    - Button: 11. Recuperar ocupació

Nota: Les dates que hi figuren s'han escollit per fer la demostració del funcionament del programa. S'hauran de substituir per les dates reals d'arrencada del servei, de dia actual (avui), etc.

- **Posada en producció:** indica els processos a executar al final del dia anterior a l'arrencada del servei (o de matinada del dia d'arrencada, un cop es disposi de tot el registre d'usuaris del dia anterior).
  - En aquest exemple, se suposa que el dia de posada en producció és el 14/07/16. S'ha fet així perquè disposàvem del registre d'usuaris dels dos mesos anteriors a aquesta data. S'haurà de substituir aquesta data per la data real d'arrencada del servei.
  - Cal fixar una data d'inici per a l'anàlisi de les dades, que haurà de correspondre aproximadament a dos mesos abans del dia d'arrencada. Ha de ser un dia pel que hi hagi registre, és a dir, no pot ser un dissabte, diumenge, ni festiu. Així, en l'exemple s'ha fixat el 13/05/16, atès que és un divendres (el dia 15/05/16 és diumenge, invàlid).
  - Per tant, en l'exemple, el període inicial d'anàlisi de la sèrie temporal (ST) està format pels dos mesos aproximats entre el 13/05/16 i el 14/07/16.
    - El període històric de la sèrie temporal, per al què es disposa de tot el registre d'usuaris, és: 13/05/16 – 13/07/16. L'estimació del model es farà en base a les dades d'aquest període. No és necessari que siguin dos mesos exactes, però sí que no siguin menys de dos mesos<sup>1</sup>.
    - El dia 14/07/16 és el primer dia per al què es volen calcular prediccions. El registre d'usuaris d'aquest dia s'anirà alimentant a mida que vagin passant els minuts. Atès que en realitat es disposa del registre de tot el dia 14/07/16, la prova simula una situació real. Així, a les 8:00 el dia 14/07/16 apareixerà en blanc, a les 8:05 ja tindrà la dada d'ocupació de les 8:00, a les 8:10 tindrà les dades d'ocupació de les 8:00 i les 8:05, etc.
    - Les prediccions es calculen tant pel dia actual com per l'endemà, que seria el 15/07/16.
  - Botó "Càlcul hora estimada de sortida". Omple la columna FHSTOP\_NOVA del full DADES\_TOT, per al període històric 13/05/16-13/07/16. Ara cada registre té una hora d'entrada i una hora de sortida vàlides.
  - Botó "Càlcul ST ocupació". Genera la ST de places ocupades segons el registre, per al període 13/05/16-13/07/16. Aquesta ST té una periodicitat de 5 minuts. No ha de contenir dades de dissabtes. També genera les variables explicatives que es faran servir en el model (regressors): dia de la setmana, franja horària, interaccions, etc.
- **De matinada diàriament:** processos a executar cada dia abans de l'hora d'inici d'operació de l'Àrea DUM (8:00).
  - Cal introduir la data del dia actual (avui). En l'exemple, suposem que la data actual és 15/07/16. És a dir, ja ha passat un dia des de l'arrencada del servei.

---

<sup>1</sup> Idealment, es treballaria amb una ST més llarga. Tanmateix, la longitud de la ST té molta repercussió sobre el temps que necessita l'algoritme per córrer. Per tant, inicialment la limitem a dos mesos. Més endavant, un cop fetes les proves de càrrega del servidor, es podrà valorar si es pot allargar la ST.



- Automàticament, es recalcula la “Data d’inici de la sèrie analitzada”. Aquesta data correspondrà aproximadament a dos mesos abans del dia actual. A mida que avancen els minuts (de 5 en 5), el període històric també es retalla pel principi. Per exemple, inicialment el període històric és 13/05/16-13/07/16; a l’endemà serà 14/05/16-14/07/16. És a dir, sempre es faran servir els dos mesos anteriors a la data i hora actual per fer l’estimació del model. Per exemple, si el dia actual és el 15/07/16, i l’hora actual són les 10:10, el període històric anirà des de les 10:10 del dia 14/05/16 fins a les 10:05 del dia 15/07/16.
  - Botó “Càlcul hora estimada de sortida”. Actualitza l’hora FHSTOP\_NOVA, que s’haurà anat generant el dia anterior a mida que hagin entrat nous registres d’usuari. Cal fer aquesta actualització diàriament, per corregir els casos en què un vehicle va romandre estacionat un temps superior a la mitjana que s’estima que hi són els vehicles que no registren la sortida.
  - Botó “Càlcul ST ocupació”. Actualitza la ST d’ocupació amb les dades del darrer dia. Calcula els regressors associats a les noves observacions, a partir de la variable que recull la data i hora.
  - Botó “Càlcul predicció AR(0)”. Calcula les prediccions segons un model AR(0), equivalent a un model de regressió lineal múltiple. Aquestes prediccions es mantenen constants durant tot el dia actual, així que només cal calcular-les una vegada al dia. A petició de BSM, aquestes prediccions s’estenen també al dia següent a l’actual.
  - Botó “Resultat”. Escriu les prediccions del model AR(0) en la columna “<8:00” del triangle de prediccions que després s’explicarà (vegeu Figura 3).
- **Cada 5 minuts:** processos a executar cada 5 minuts (o interval de temps mínim que sigui possible segons els recursos de càrrega de dades i càlcul disponibles), entre les 8:00 i les 19:55.
    - Hora actual: hora a la què es vol fer la predicció. L’exemple mostra el cas de les 10:10, però la predicció es fa cada 5 minuts, començant a les 8:00 i acabant a les 19:55.
    - Botó “Càlcul hora estimada de sortida”. Es recalcula l’hora estimada de sortida dels registres del dia actual entre les 8:00 i les 10:05 (en l’exemple de la Figura 2). És necessari anar fent aquest recàlcul perquè es poden haver imputat erròniament hores de sortida a vehicles que en realitat no havien sortit (pensant que no havien registrat la sortida), i més tard sí que han registrat la sortida.
    - Botó “Càlcul ST ocupació”. Calcula la ST d’ocupació de periodicitat 5 minuts per al període del dia actual transcorregut, tenint en compte els registres modificats al pas anterior.
    - Botó “Càlcul predicció AR(1)”. L’algorisme calcula el vector de prediccions corresponent a l’hora actual que acabem d’esmentar. Són prediccions per a la resta del dia i per a l’endemà, cada 5 minuts. Com s’explicarà després, el model AR(1) només s’utilitza per fer les prediccions de la primera hora. Després, les prediccions són les que s’han calculat un cop al dia amb el model AR(0).



- Botó “Resultat”. Escriu en el triangle (Figura 3) el vector de prediccions, conjugant el model AR(0) i el model AR(1) com s’ha especificat.
- **Petició de predicció d’ocupació:** simula la petició d’un usuari d’Àrea DUM.
    - L’usuari vol estacionar a la zona DUM que s’especifica. En l’exemple, és la zona 1678, ara són les 10:10, i vol estacionar avui a les 10:40.
    - Recordem que la zona DUM ha de pertànyer al subàmbit 1 o 14.
    - Dia de predicció: avui o demà, segons es vulgui la predicció per una hora del dia actual, o del dia següent.
    - Hora de predicció: hora per la què l’usuari fa la petició de previsió d’ocupació. Ha d’estar entre les 8:00 i les 19:55, i ha de ser estrictament posterior a l’hora actual.
    - Ocupació registrada: output de l’algoritme per al subàmbit i hora demanats. És la predicció que s’ha calculat amb el botó “Resultat”, a executar cada 5 minuts. Prové de la casella de la Figura 3 que creua l’hora actual amb l’hora que es vol predir.
    - Ocupació total: correcció del biaix de no registre. A l’ocupació registrada, se li afegeix un percentatge de les places operatives on s’estima que s’ha estacionat sense registre.
    - Percentatge ocupació total: es calcula quin és el percentatge de les places operatives que es troben ocupades. Aquest resultat ha de donar lloc al semàfor verd, groc o vermell, segons el criteri que BSM desitgi adoptar per a cada color.
    - Botó “Recuperar ocupació”. Executa els càlculs anteriors.
- c) Pestanya CONFIGURACIÓ. Conté diferents elements necessaris per al funcionament de l’algoritme:
- Llistat d’hores i dies (avui o demà), que la interfície permet escollir.
  - Taula amb les zones DUM contingudes a la zona pilot. Per a cada zona, conté el nombre de places operatives (caldrà actualitzar-les cada dia amb la informació que BSM proporciona), l’àmbit i subàmbit a què pertany la zona DUM, i nombre de places operatives de tot el subàmbit (igualment, s’actualitzarà cada dia).
  - Taula per àmbit i franja horària. Els percentatges per als 4 àmbits corresponen als percentatges de les places operatives del subàmbit que s’estima que s’ocupen sense registre (coincideix amb la Taula 7.4. en apartat posterior). La filera de xifres amb dos decimals correspon al temps mitjà d’estacionament que s’imputa als registres sense FHSTOP vàlida (coincideix amb valors de la Taula 3). Aquests paràmetres s’aniran actualitzant a mida que es disposi de feedback real sobre la fiabilitat de les prediccions d’ocupació. Igualment, s’hauran de revisar quan el nivell de registre dels usuaris augmenti.
  - Minuts canvi de model. Es tracte d’un paràmetre tècnic que no s’ha de modificar si no ho indica l’equip d’IREA-UB. Correspon als minuts que es tarda en passar de la predicció segons el model AR(1) a la predicció segons el model AR(0). Inicialment, suposarem que són 60 minuts.

- Path RScript: ruta on es troba el fitxer RScript.exe, que forma part del programa R, i que es necessita per cridar l'executable de R que fa les estimacions i les prediccions dels models.
  - Path Dades: ruta on es troba el fitxer Excel amb el programa en VBA. En aquest directori es generaran automàticament alguns fitxers necessaris per als càlculs del programa.
- d) Pestanya Dades5min: conté la ST d'ocupacions registrades per als dos mesos d'anàlisi, amb periodicitats de 5 minuts. Inclou també els regressors associats a cada observació. La darrera columna conté un filtre que el programa VBA anirà calculant segons quin sigui el període per fer la predicció cada dia i cada hora.
- e) Pestanya Res5min\_1. Conté les prediccions que es vagin generant per al subàmbit 1. Vegeu Figura 3 per a un subàmbit genèric. Per exemple, si s'executa el programa amb hora actual 10:10, omplirà la columna de les 10:10 amb les prediccions a partir d'aquesta hora i per a l'endemà. Les prediccions per a l'endemà apareixen en vermell.
- f) Pestanya Res5min\_14. Conté les prediccions que es vagin generant per al subàmbit 14. Vegeu Figura 3 per a un subàmbit genèric. Igual que l'anterior.

L'apartat següent detalla amb més extensió quin és el flux de l'algoritme VBA.

#### 4. Flux de l'algoritme VBA

L'algoritme VBA segueix el següent flux. Les 9 etapes es detallen a continuació:

1. Entrada al sistema de les dades de registre d'usuaris DUM.
2. Generació d'hora de sortida als registres amb hora de sortida inexistent o no vàlida.
3. Generació de la sèrie temporal d'ocupació registrada per a cada subàmbit.
4. Càlcul de les variables associades a cada valor de la sèrie temporal d'ocupació (regressors).
5. Estimació dels models de sèrie temporal.
6. Prediccions de l'ocupació registrada en base als models de sèrie temporal.
7. Correcció de les prediccions per introduir l'ocupació no registrada.
8. Càlcul del percentatge d'ocupació.
9. Transformació del percentatge d'ocupació en semàfor verd, groc o vermell.

##### 4.1. Entrada al sistema de les dades de registre d'usuaris DUM

Els usuaris de vehicles comercials que estacionen a un àrea DUM tenen obligació d'enregistrar-se a través de SMS o de l'aplicatiu dissenyat a l'efecte. Malgrat aquesta obligació, un elevat nombre d'usuaris no s'enregistren.

Cada usuari que es registra genera informació sobre diverses variables. Les d'interès per a l'algoritme són:

- FHSTART: data i hora de registre, inici de l'estacionament.
- FHSTOP: data i hora de sortida de l'estacionament.
- ID\_ZONADUM: identificador de la zona DUM en què l'usuari estaciona.

Aquestes dades formen un flux continu de registres on van apareixent nous casos cada dia entre les 8:00 i les 20:00, a mida que els usuaris fan servir l'aplicació de registre. Com veurem, aquests registres tenen dues limitacions fonamentals:

- Representen només un **subconjunt dels usuaris que estacionen**, atès que hi ha un nombre molt significatiu d'usuaris que no s'enregistren. Es calcula que al voltant del **45%** de les places operatives són ocupades per usuaris no registrats. No es disposa de cap informació sobre aquests usuaris: ni registre d'entrada, ni registre de sortida.
- Contenen **informació parcial**: el registre de sortida fins al moment no és obligatori. Hi ha un nombre molt significatiu d'usuaris que registren l'arribada a l'estacionament, però no la sortida (FHSTOP no informada). Molts d'aquests casos són completats automàticament pel sistema, que assigna final de l'estacionament als 30 minuts de l'arribada. Aquests valors de sortida es consideraran no vàlids. S'estima que al voltant del **61%** dels registres tenen FHSTOP no informat o invàlid.

**Taula 2.** Extracte del registre d'usuaris DUM (exemple).

Registre d'usuaris DUM			Observacions
FHSTART	FHSTOP	ID_ZONADUM	
06/07/2016 08:00:37	06/07/2016 08:10:02	1151	Correcte
06/07/2016 08:01:00	06/07/2016 08:22:29	1678	Correcte
06/07/2016 08:01:07	06/07/2016 08:31:07	1047	No vàlid: 30 minuts
06/07/2016 08:01:22	06/07/2016 08:31:22	1333	No vàlid: 30 minuts
06/07/2016 08:01:49		1459	No vàlid: FHSTOP no informat
06/07/2016 08:02:12	06/07/2016 08:53:31	1365	Correcte
06/07/2016 08:03:33	06/07/2016 08:23:41	1485	Correcte
06/07/2016 08:03:38	06/07/2016 08:33:38	1046	No vàlid: 30 minuts
06/07/2016 08:04:06	06/07/2016 08:20:07	1485	Correcte
06/07/2016 08:06:51	06/07/2016 08:32:04	1047	Correcte
06/07/2016 08:07:26	06/07/2016 08:37:26	1675	No vàlid: 30 minuts
06/07/2016 08:07:57	06/07/2016 08:19:17	1048	Correcte

#### 4.2. Generació d'hora de sortida als registres amb hora de sortida inexistent o no vàlida

Els casos en què l'hora de sortida (FHSTOP) no es considera vàlida són els següents:

- FHSTOP no informada. El camp apareix en blanc.
- FHSTOP correspon a 30 minuts exactes després de FHSTART.
- FHSTOP és igual a 8:30:00 i FHSTART és anterior a les 8:00.
- El dia d'inici d'estacionament (FHSTART) i el de finalització (FHSTOP) no coincideixen.

Com s'ha comentat, els casos que es troben en alguna d'aquestes situacions representen aproximadament un 61% del total dels registres. La imputació d'una hora de sortida és necessària per a la construcció de la sèrie temporal de vehicles comercials aparcats en cada moment.

Per tal de reunir informació sobre el temps d'estada dels vehicles i poder fer la imputació de FHSTOP en els casos no vàlids, BSM va dur a terme un treball de camp (TdC) de **durada de l'estada** a la zona pilot. Aquest es va desenvolupar durant les setmanes del 28/11/16 al 3/12/16 (àmbits 1 i 2), i 12/12/16 al 17/12/16 (àmbits 3 i 4). En ell es va obtenir la durada de l'estacionament de 4654 vehicles comercials. L'anàlisi d'aquestes dades mostra que existeixen algunes diferències per franja horària. A partir d'aquesta informació i del càlcul de la durada mitjana per als casos amb FHSTOP vàlid (casos informats), s'ha calculat quina és la durada de l'estacionament més versemblant per als registres amb FHSTOP no informat.

**Taula 3.** Estada mitjana per franja horària, en minuts.

Estada mitjana	Franja horària					
	8-10	10-12	12-14	14-16	16-18	18-20
Segons el TdC	36,88	33,65	31,56	40,29	30,78	21,95
Casos informats	25,37	23,99	24,01	24,33	19,92	19,32
Casos no informats	45,37	40,10	36,14	47,96	37,02	23,62

**Taula 4.** Imputació de FHSTOP\_NOVA.

Registre d'usuari DUM			
FHSTART	FHSTOP	ID_ZONADUM	FHSTOP_NOVA
06/07/2016 08:00:37	06/07/2016 08:10:02	1151	06/07/2016 08:10:02
06/07/2016 08:01:00	06/07/2016 08:22:29	1678	06/07/2016 08:22:29
06/07/2016 08:01:07	06/07/2016 08:31:07	1047	06/07/2016 08:46:29
06/07/2016 08:01:22	06/07/2016 08:31:22	1333	06/07/2016 08:46:44
06/07/2016 08:01:49		1459	06/07/2016 08:47:11
06/07/2016 08:02:12	06/07/2016 08:53:31	1365	06/07/2016 08:53:31
06/07/2016 08:03:33	06/07/2016 08:23:41	1485	06/07/2016 08:23:41
06/07/2016 08:03:38	06/07/2016 08:33:38	1046	06/07/2016 08:49:00
06/07/2016 08:04:06	06/07/2016 08:20:07	1485	06/07/2016 08:20:07
06/07/2016 08:06:51	06/07/2016 08:32:04	1047	06/07/2016 08:32:04
06/07/2016 08:07:26	06/07/2016 08:37:26	1675	06/07/2016 08:52:48
06/07/2016 08:07:57	06/07/2016 08:19:17	1048	06/07/2016 08:19:17

Els valors de la darrera filera de la Taula 3 ja es troben incorporats al mòdul corresponent de l'algoritme VBA, que assigna una FHSTOP vàlida (anomenada FHSTOP\_NOVA) als registres que tenien FHSTOP no vàlida. Es fa amb els botons "Càlcul hora estimada de sortida".

Cal tenir present que, a mida que es vagin generant registres, caldrà anar completant la FHSTOP\_NOVA en aquests casos:

- El sistema assigna FHSTOP passats 30 minuts exactes de FHSTART (es considera FHSTOP no vàlid, atès que no ha estat introduït per l'usuari).
- El camp FHSTOP segueix en blanc passats els minuts recollits a la darrera filera de la taula 3 (segons la franja horària).

En aquests dos casos, atès que no se sap si el vehicle ha marxat o no, se suposarà que ha marxat en els minuts calculats per als casos no informats (Taula 3). Això es duu a terme amb el botó de "Càlcul d'hora estimada de sortida" de cada 5 minuts.

Posteriorment, cada nit caldrà actualitzar la FHSTOP\_NOVA, perquè hi haurà un petit nombre de casos en què l'usuari registrarà que marxa més tard de l'hora FHSTOP\_NOVA que se li ha assignat (s'ha calculat que això passa en un 3% dels casos, amb diferències significatives amb l'hora assignada en un 1% dels casos). Aquest fet pràcticament no té conseqüències a temps real, ja que només suposa un petit desajust en la sèrie temporal tenint en compte que aquesta té una longitud de 2 mesos, però igualment es corregeix. Tanmateix, cada nit es fa el recàlcul, de forma que la sèrie temporal històrica sigui el més fidel possible a l'actualitat.

### **4.3. Generació de la sèrie temporal d'ocupació registrada per a cada subàmbit**


La predicció de l'ocupació en cada moment es basa en la sèrie històrica de registre d'ocupació. Aquesta sèrie històrica és una sèrie temporal (ST), que es pot generar amb diferents periodicitats. Per tant, s'ha optat inicialment per construir la ST de registres amb una periodicitat de 5 minuts. L'algoritme VBA calcula automàticament aquesta ST.

Per tal de visualitzar què significa això, imaginem que fem una foto del subàmbit cada 5 minuts, i mirem quants vehicles hi ha estacionats (en realitat, veurem només els que s'han enregistrat, ja que per a la resta no tenim dades). Així, per a cada subàmbit DUM tindrem 144 observacions diàries (8:00; 8:05; 8:10,..., 19:55). Vegeu columna ST, de la pestanya "Dades5 min".

Suposem, per exemple, que els registres de la Taula 5 corresponen al subàmbit 1 (ID\_ZONADUM=1478 i ID\_ZONADUM=1487). El programa VBA compta quants vehicles hi ha estacionats en la foto de les 8:00, en la foto de les 8:05, en la de les 8:10, etc.

**Taula 5.** Generació de la sèrie temporal d'ocupació registrada (periodicitat 5 minuts).

FHSTART	FHSTOP_NOVA	SUBÀMBIT
06/07/2016 08:00:37	06/07/2016 08:10:02	1
06/07/2016 08:01:00	06/07/2016 08:22:29	1
06/07/2016 08:01:07	06/07/2016 08:46:29	1
06/07/2016 08:01:22	06/07/2016 08:46:44	1
06/07/2016 08:01:49	06/07/2016 08:47:11	1
06/07/2016 08:02:12	06/07/2016 08:53:31	1
06/07/2016 08:03:33	06/07/2016 08:23:41	1
06/07/2016 08:03:38	06/07/2016 08:49:00	1
06/07/2016 08:04:06	06/07/2016 08:20:07	1
06/07/2016 08:06:51	06/07/2016 08:32:04	1
06/07/2016 08:07:26	06/07/2016 08:52:48	1
06/07/2016 08:07:57	06/07/2016 08:19:17	1



Hora	Sèrie temporal: nre. de vehicles registrats
8:00	0
8:05	9
8:10	12
8:15	11
8:20	10
8:25	7
8:30	7
...	...

Algunes observacions:

- A efectes de programació, es considera que un vehicle està estacionat a les hh:mm:ss, si ho està entre les hh:mm:00 i les hh:mm:59. Per exemple, es considera que està estacionat a les 8:00 si ho està entre les 8:00:00 i les 8:00:59.
- Les ST es calcularan per a cada **subàmbit**, atès que aquesta serà la unitat territorial per a la que es donaran les prediccions.
- La longitud en dies de les sèries temporals dels subàmbits és parametrizable. Com s'ha dit, i fins que es pugui testar la capacitat del servidor, es proposa que les sèries tinguin una longitud de dos mesos. A mida que es vagi incorporant un nou valor de la ST (cada 5 minuts), s'eliminarà un valor de l'inici de la sèrie. Així, la ST tindrà la mateixa longitud en tot moment.
- Els dissabtes s'han exclòs de l'anàlisi durant aquesta primera fase de l'estudi. Per tant, dissabtes, diumenges i festius no formen part de la ST d'ocupació registrada.
- S'ha fixat la periodicitat en 5 minuts atès que, a menor temps entre la darrera dada disponible i la predicció que es vol fer, major serà la fiabilitat d'aquesta darrera. Cas que, per motius operatius de BSM, no sigui possible disposar de l'ocupació de fa 5 minuts (o els càlculs que es derivin suposin excessiva càrrega per al servidor), caldrà ampliar aquest interval de temps. Sempre es farà la mínima ampliació possible, entenent, per exemple, que 10 minuts és preferible a 15 minuts.
- Per a cada dia, la sèrie temporal té 144 observacions: de 8:00 a 19:55. Les 20:00 s'exclouen. Si un usuari demana la predicció d'ocupació per a les 20:00 s'ha de mostrar un missatge indicant que l'àrea DUM ja no és operativa a aquesta hora (o bé, cas que l'usuari hagi de seleccionar l'hora per a la que vol la predicció, es pot impedir que seleccioni les 20:00).
- En el dia actual, cada 5 minuts s'ha de generar una nova observació de la sèrie temporal. Això ho fa el botó "Càlcul ST ocupació" cada 5 minuts.

#### 4.4. Càlcul de les variables associades a la ST d'ocupació (regressors)

La ST generada proporciona la informació històrica necessària per predir l'ocupació. La predicció es farà en base a models on hi ha una variable dependent (la pròpia ST, és a dir, el nombre de places ocupades segons el registre), que es vol explicar a partir d'altres variables, que s'anomenen regressors. Els regressors que s'incorporen en aquesta fase del projecte són: el dia de la setmana, la franja horària, algunes interaccions entre elles, i la setmana del mes.

Per tal de poder incloure els regressors en el model, és necessari definir-los com a variables binàries. Per exemple, el regressor DM (dimarts) valdrà 1 si la data de registre correspon a un dimarts, i 0 en cas contrari. És important tenir present que cal deixar un dia de la setmana (i una franja horària) com a dia/franja de "referència", no incloent-los en el model. Altrament, hi haurà un problema de colinealitat perfecta i no serà possible realitzar l'estimació del model. Per tant, els regressors que inclouran els models seran:

- Variables de dia de la setmana. DM: dimarts (1=sí; 0=no); DC: dimecres (1=sí; 0=no); DJ: dijous (1=sí; 0=no); DV: divendres (1=sí; 0=no). Dilluns serà el dia de referència i no s'inclourà. Igualment, recordem que dissabtes i diumenges no formen part de l'estudi. Són 4 regressors.
- Variables de franja horària. Són diferents per a les prediccions basades en els models AR(0) i AR(1). El motiu és que el model AR(0) només s'haurà d'executar un cop al dia per a cada subàmbit, de forma que el major nombre de regressors del model no sembla un problema des del punt de vista del consum de recursos del servidor. En canvi, el model AR(1) s'haurà d'executar cada 5 minuts, de manera que és preferible minimitzar el nombre de regressors que inclou.
  - Model AR(0): franges horàries de 30 minuts, de 8:00:00 a 19:29:59. FH8-8': franja horària de 8:00:00 a 8:29:59 (1=sí; 0=no); FH8'-9: franja horària de 8:30:00 a 8:59:59; ...; FH19-19': franja horària de 19:00:00 a 19:29:59 (1=sí; 0=no). La franja horària de 19:30:00 a 19:59:59 serà la franja horària de referència i no s'inclourà en el model. Són 23 regressors.
  - Model AR(1): franges horàries d'una hora, de 8:00:00 a 18:59:59. FH8-9: franja horària de 8:00:00 a 8:59:59 (1=sí; 0=no); ...; FH18-19: franja horària de 18:00:00 a 18:59:59 (1=sí; 0=no). La franja horària de 19:00:00 a 19:59:59 serà la franja horària de referència i no s'inclourà en el model. Són 11 regressors.
- Variables d'interacció entre dia de la setmana i franja horària. Atès que els divendres a la tarda semblen tenir un comportament força diferenciat de la resta, s'inclouen les variables: DV14-15: divendres de 14:00:00 a 14:59:59 (1=sí; 0=no); ...; DV19-20: divendres de 19:00:00 a 19:59:59 (1=sí; 0=no). Són 6 regressors.
- Variables de setmana del mes. SET2: dia del mes  $\geq 8$  i  $< 17$  (1=sí; 0=no); SET3: dia del mes  $\geq 17$  i  $< 26$  (1=sí; 0=no); SET4: dia del mes  $\geq 26$ . La variable SET1, dia del mes  $< 8$ , no s'inclou en el model atès que actua com a variable de referència. Són 3 regressors.



Taula 6. Regressors associats a algunes observacions de la sèrie temporal d'ocupació registrada.\*

Data i hora	SUBÀMBIT	ST	DM	DC	DJ	DV	DV14-15	DV15-16	DV16-17	DV17-18	DV18-19	DV19-20	FH8-9	FH9-10	FH10-11	FH11-12	FH12-13	FH13-14	FH14-15	FH15-16	FH16-17	FH17-18	FH18-19	SET2	SET3	SET4	
25/05/2016 19:00	1	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
25/05/2016 19:05	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:10	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:15	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:20	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:25	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:30	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:35	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:40	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:45	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:50	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25/05/2016 19:55	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
26/05/2016 08:00	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:05	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:10	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:15	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:20	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:25	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:30	1	2	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:35	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:40	1	2	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:45	1	2	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:50	1	3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 08:55	1	3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 09:00	1	4	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
26/05/2016 09:05	1	5	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1

\*El model AR(0) utilitza franges horaries de 30 minuts, enlloc d'una hora.

#### 4.5. Estimació dels models de sèrie temporal

Amb l'objectiu de predir l'ocupació registrada, és necessari construir models econòmics que expliquin el nombre de vehicles estacionats en cada moment, en funció dels regressors mostrats a l'apartat anterior (dia de la setmana, franja horària, setmana del mes, i interaccions).

Els models indicats per explicar una sèrie temporal són els models estocàstics ARIMA (Autoregressive Integrated Moving Average models)<sup>2</sup>. Aquests models expliquen el valor de la variable en el moment  $t$ , en funció dels valors en els moments  $t-1$ ,  $t-2$ ,...,  $t-k$ . El període entre una observació i la següent és, com ja s'ha dit, de 5 minuts.

En aquest cas, s'ha escollit un model AR(1), autorregressiu d'ordre 1. Aquest model prediu el valor de la sèrie en el moment  $t$ ,  $Y_t$ , en funció de:

- el valor de la sèrie en el moment  $t-1$  ( $Y_{t-1}$ ),
- i els regressors, que indiquen quina és la franja horària, el dia de la setmana, etc.

Tanmateix, després de nombroses proves, s'ha conclòs que la memòria del procés és de 60 minuts. Això significa que, passats 60 minuts, el valor de  $Y_{t-1}$  no aporta informació útil per predir el valor de  $Y_t$ . Per tant:

- Es farà servir el model AR(1) quan el desfasament entre l'hora actual (hora a la que es rep la petició de l'usuari) i l'hora prevista d'estacionament (hora a la que l'usuari vol estacionar) sigui de 60 minuts o inferior.
- En cas contrari, el model no necessita incorporar el valor de  $Y_{t-1}$ . Aleshores, es farà servir un model de regressió lineal múltiple (MRLM), que només incorpora com a variables explicatives els regressors abans citats. Aquest model és equivalent a un AR(0), autorregressiu sense cap retard com a variable explicativa. A efectes de claredat de programació, el MRLM s'implementarà com un AR(0).
- Igualment, si l'usuari fa la petició abans de les 8:00, o abans del moment en què s'hagi calculat (en base al registre) la primera ocupació del dia, el model utilitzat serà l'AR(0), atès que no existeix cap valor de  $Y_{t-1}$  corresponent al mateix dia.

L'estimació dels models ARIMA requereix la utilització d'un software estadístic. En l'actualitat el software més utilitzat és el programa R, atès que implementa la metodologia estadística i econòmica més recent, i es pot descarregar gratuïtament a [www.r-project.org](http://www.r-project.org).

VBA i R interactuen de forma senzilla, a través d'un fitxer R executable. Així:

- VBA duu a terme les primeres etapes 1-4 del procés.
- VBA genera un fitxer ".csv" que conté la sèrie temporal d'ocupació registrada (etapa 3) i els regressors (etapa 4).

<sup>2</sup> Per a la descripció detallada de la metodologia ARIMA podeu consultar: Ashley, R.A. (2012) Fundamentals of Applied Econometrics. John Wiley & Sons, Inc.

- A través d'un fitxer R executable (integrat en el programa VBA), R llegeix aquestes dades, crida les llibreries necessàries, estima els models AR(0) i AR(1) i fa les prediccions de l'ocupació.
- R genera un fitxer “.csv” amb l'output, concretament amb les prediccions.
- VBA recupera aquest fitxer, i a partir d'aquí es segueix amb les següents etapes del flux amb VBA.

Aquest procés es repeteix cada 5 minuts, comptant que el registre d'usuaris s'alimenta de forma contínua. Així, cada 5 minuts s'haurà afegit una nova observació de la sèrie temporal i s'hauran calculat els regressors associats a aquesta nova observació.

#### 4.6. Prediccions de l'ocupació registrada en base als models de sèrie temporal

Un cop estimats els models AR(0) i AR(1), cal procedir a la predicció de l'ocupació registrada. Aquestes prediccions les duu a terme el fitxer R executable (integrat en l'algoritme VBA), utilitzant un algoritme conegut com Kalman Forecast, que millora de forma significativa la fiabilitat de les prediccions. Les prediccions s'actualitzen cada 5 minuts, un cop s'ha reestimat el model amb l'ocupació registrada del darrer període disponible.

La Figura 3 mostra un exemple de la seqüència de prediccions que es fan al llarg d'un dia:

- a) De matinada s'executa el model AR(0), colorejat en verd.
  - Les prediccions basades en aquest model s'utilitzen: (i) per a peticions d'ocupació realitzades abans de les 8:00; (ii) per a prediccions realitzades a qualsevol hora, en el cas que hi hagi un desfasament de més de 60 minuts entre l'hora actual (hora de la petició) i l'hora per a la que l'usuari vol conèixer l'ocupació.
  - Aquestes prediccions no s'actualitzen al llarg del dia, atès que no tenen en compte el valor d'ocupació dels moments anteriors (període t-1). Per exemple, la predicció per a les 10:00 és la mateixa si la petició s'ha fet abans de les 8:00, que si s'ha fet a les 8:55 (valor p0\_25 a la Taula 6). A partir de les 9:00, la predicció serà la del model AR(1), perquè ja no hi ha més d'una hora entre l'hora de la petició (les 9:00) i l'hora a predir (les 10:00).
  - L'estimació i predicció del model AR(0) s'haurà d'actualitzar cada matinada, un cop s'hagin incorporat a la sèrie històrica els valors d'ocupació de la sèrie temporal del dia anterior, que ja haurà finalitzat.
- b) A partir de les 8:00, s'executa el model AR(1) cada 5 minuts (valors colorejats en groc). Les prediccions van evolucionant a mida que entren noves observacions (generades pel registre) a la sèrie temporal.
  - Per exemple, la predicció per a les 9:00 és diferent si l'hora actual (hora de la petició) és les 8:00 (p1\_12), les 8:05 (p2\_11), les 8:10 (p3\_10), etc. La idea és que aquesta predicció es va afinant més a mida que s'hi acosten les 9:00.

Figura 3. Extracte de les prediccions d'un dia per a un àmbit, segons el model utilitzat.

		Hora actual (l'usuari fa la petició)													
		<8:00	08:00	08:05	08:10	08:15	08:20	08:25	08:30	08:35	08:40	08:45	08:50	08:55	09:00
Hora a la què es vol predir l'ocupació	08:00	p0_1													
	08:05	p0_2	p1_1												
	08:10	p0_3	p1_2	p2_1											
	08:15	p0_4	p1_3	p2_2	p3_1										
	08:20	p0_5	p1_4	p2_3	p3_2	p4_1									
	08:25	p0_6	p1_5	p2_4	p3_3	p4_2	p5_1								
	08:30	p0_7	p1_6	p2_5	p3_4	p4_3	p5_2	p6_1							
	08:35	p0_8	p1_7	p2_6	p3_5	p4_4	p5_3	p6_2	p7_1						
	08:40	p0_9	p1_8	p2_7	p3_6	p4_5	p5_4	p6_3	p7_2	p8_1					
	08:45	p0_10	p1_9	p2_8	p3_7	p4_6	p5_5	p6_4	p7_3	p8_2	p9_1				
	08:50	p0_11	p1_10	p2_9	p3_8	p4_7	p5_6	p6_5	p7_4	p8_3	p9_2	p10_1			
	08:55	p0_12	p1_11	p2_10	p3_9	p4_8	p5_7	p6_6	p7_5	p8_4	p9_3	p10_2	p11_1		
	09:00	p0_13	p1_12	p2_11	p3_10	p4_9	p5_8	p6_7	p7_6	p8_5	p9_4	p10_3	p11_2	p12_1	
	09:05	p0_14	p0_14	p2_12	p3_11	p4_10	p5_9	p6_8	p7_7	p8_6	p9_5	p10_4	p11_3	p12_2	p13_1
	09:10	p0_15	p0_15	p0_15	p3_12	p4_11	p5_10	p6_9	p7_8	p8_7	p9_6	p10_5	p11_4	p12_3	p13_2
	09:15	p0_16	p0_16	p0_16	p0_16	p4_12	p5_11	p6_10	p7_9	p8_8	p9_7	p10_6	p11_5	p12_4	p13_3
	09:20	p0_17	p0_17	p0_17	p0_17	p0_17	p5_12	p6_11	p7_10	p8_9	p9_8	p10_7	p11_6	p12_5	p13_4
	09:25	p0_18	p0_18	p0_18	p0_18	p0_18	p0_18	p6_12	p7_11	p8_10	p9_9	p10_8	p11_7	p12_6	p13_5
	09:30	p0_19	p0_19	p0_19	p0_19	p0_19	p0_19	p0_19	p7_12	p8_11	p9_10	p10_9	p11_8	p12_7	p13_6
	09:35	p0_20	p0_20	p0_20	p0_20	p0_20	p0_20	p0_20	p0_20	p8_12	p9_11	p10_10	p11_9	p12_8	p13_7
	09:40	p0_21	p0_21	p0_21	p0_21	p0_21	p0_21	p0_21	p0_21	p0_21	p9_12	p10_11	p11_10	p12_9	p13_8
	09:45	p0_22	p0_22	p0_22	p0_22	p0_22	p0_22	p0_22	p0_22	p0_22	p0_22	p10_12	p11_11	p12_10	p13_9
	09:50	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p0_23	p11_12	p12_11	p13_10
	09:55	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p0_24	p12_12	p13_11
	10:00	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p0_25	p13_12

Prediccions AR(0)  
Prediccions AR(1)

#### 4.7. Correcció de les prediccions per introduir l'ocupació no registrada

Recordem que fins ara s'ha predit l'ocupació registrada, ja que les dades que han entrat en el model són les provinents del registre d'usuaris d'Àrea DUM. Com ja s'ha comentat, no tots els usuaris s'enregistren quan estacionen, malgrat l'obligatorietat del registre per als vehicles comercials (no, per exemple, per als vehicles de minusvàlids). Aquest fet dona lloc a un important biaix entre l'ocupació registrada i l'ocupació real.

Per tal d'estimar aquest biaix, BSM va dur a terme un treball del camp d'ocupació entre el 13/06/16 i el 17/16/16 (TdC ocupació). En ell, un equip d'agents anotava al llarg de tot el dia quants vehicles hi havia estacionats en les àrees DUM de la zona pilot, junt amb la seva tipologia. Els agents passaven per cada àrea amb una freqüència aproximada de 15-20 minuts. Els resultats es mostren a la Taula 7.1.

Adicionalment, es van recollir algunes dades més d'ocupació real a través dels vigilants d'Àrea, que van anotar l'ocupació a la zona pilot mentre realitzaven les seves tasques ordinàries de vigilància entre el 07/11/106 i el 17/12/16 (TdC vigilants). Els resultats es mostren a la Taula 7.2.

D'aquesta manera, s'ha reunit un volum suficient d'informació sobre ocupació real. Aquests resultats s'han comparat amb l'ocupació registrada (Taula 7.3.), i s'ha estimat el biaix que hi ha entre l'ocupació registrada i l'ocupació real (Taula 7.4). Els càlculs s'han desagregat per franja horària i per àmbit, atès que s'observen algunes diferències que cal tenir en compte.

**Taula 7.1.** Percentatge de places operatives ocupades, segons el TdC d'ocupació

Àmbit	Franja horària						TOTAL
	8-10	10-12	12-14	14-16	16-18	18-20	
1	61%	77%	81%	83%	75%	70%	74%
2	86%	95%	90%	79%	75%	72%	83%
3	76%	91%	90%	81%	77%	72%	82%
4	40%	62%	62%	55%	55%	52%	56%
<b>Total zona pilot</b>	<b>65%</b>	<b>78%</b>	<b>79%</b>	<b>74%</b>	<b>70%</b>	<b>65%</b>	<b>72%</b>

**Taula 7.2.** Percentatge de places operatives ocupades, segons el TdC de vigilants

Àmbit	Franja horària						TOTAL
	8-10	10-12	12-14	14-16	16-18	18-20	
1	52%	68%	83%	76%	76%	75%	72%
2	85%	87%	90%	69%	79%	83%	84%
3	57%	73%	77%	80%	76%	57%	70%
4	46%	64%	67%	58%	64%	65%	60%
<b>Total zona pilot</b>	<b>57%</b>	<b>70%</b>	<b>79%</b>	<b>73%</b>	<b>74%</b>	<b>68%</b>	<b>70%</b>

**Taula 7.3.** Percentatge de places operatives amb ocupació registrada

Àmbit	Franja horària						TOTAL
	8-10	10-12	12-14	14-16	16-18	18-20	
1	20%	36%	31%	22%	23%	10%	24%
2	28%	45%	38%	30%	27%	13%	30%
3	21%	45%	41%	30%	27%	12%	29%
4	16%	35%	30%	27%	24%	12%	24%
<b>Total zona pilot</b>	<b>21%</b>	<b>40%</b>	<b>35%</b>	<b>27%</b>	<b>25%</b>	<b>11%</b>	<b>26%</b>

**Taula 7.4.** Percentatge de places operatives que s'ocupen sense registre (biaix)

Àmbit	Franja horària						TOTAL
	8-10	10-12	12-14	14-16	16-18	18-20	
1	38%	37%	50%	57%	53%	60%	50%
2	57%	50%	52%	48%	49%	60%	53%
3	47%	38%	44%	51%	50%	57%	48%
4	26%	28%	33%	28%	33%	42%	32%
<b>Total zona pilot</b>	<b>41%</b>	<b>35%</b>	<b>44%</b>	<b>47%</b>	<b>47%</b>	<b>54%</b>	<b>45%</b>

Els resultats de la Taula 7.4. són necessaris per corregir l'ocupació registrada i estimar l'ocupació real. Aquests percentatges (per franja horària i àmbit) es faran servir per incrementar la predicció que fan els models de l'ocupació. Per exemple, si per al subàmbit 9 (que pertany a l'àmbit 1, i té un total de 11 places operatives), el model ha donat una predicció de 2,4 places ocupades, i estem a la franja horària 12-14, caldrà modificar aquest valor de la predicció fent:

$$\text{Estimació de les places ocupades reals} = 2,4 + 0,50 \cdot 11 = 7,9 \text{ places}$$

Si el resultat excedís les 11 places operatives, el percentatge d'ocupació es trucarà al 100% en el pas següent.

Cal recordar que el nombre de places operatives a cada zona DUM és dinàmic i que s'haurà d'actualitzar diàriament.

#### 4.8. Càlcul del percentatge d'ocupació

Un cop dut a terme el pas anterior, el càlcul del percentatge d'ocupació real és trivial, i es fa només quan hi ha una petició d'un usuari. En el cas de l'exemple:

$$\% \text{ ocupació real} = 7,9/11 \cdot 100 = 71,8\%$$

Cas que aquest percentatge excedís el 100%, caldria truncar-lo al 100%, indicant ocupació completa del subàmbit.

#### 4.9. Transformació del percentatge d'ocupació en semàfor verd, groc o vermell

Un cop obtinguts els percentatges anteriors, BSM fixarà els intervals d'ocupació en què es considerarà, de cara a l'usuari, que hi ha places lliures amb alta probabilitat (semàfor verd), amb probabilitat moderada (semàfor groc), o amb baixa probabilitat (semàfor vermell).

### 5. Comentaris sobre línies de futur

- Les prediccions calculades per l'algoritme són susceptibles de millora. A mida que una proporció més elevada d'usuaris s'enregistren per estacionar en zona DUM, la qualitat de les dades s'incrementarà. Cal tenir present que ara mateix el percentatge d'ocupació no registrada s'ha estimat en el 45%. Un major registre implicarà menor necessitat de realitzar estimacions (usuaris no registrats, estada dels usuaris sense hora de sortida) i, per tant, major fiabilitat de les prediccions. Igualment, en el moment que BSM incorpori el sistema previst per detectar que un usuari abandona l'estacionament, les dades de registre guanyaran molta qualitat. En aquests moments, aproximadament el 61% dels registres no tenen hora de sortida vàlida. Igualment, si es decideix obligar els minusvàlids a enregistrar-se, la qualitat de les dades registrades d'ocupació creixerà significativament.
- Es treballarà per identificar i incorporar noves variables explicatives que perfilin millor les prediccions. Un cop es disposi d'una sèrie de registres més llarga serà possible incorporar, per exemple, l'efecte del mes de l'any. També s'exploraran amb més detall possibles interaccions entre dies de la setmana i franges horàries, etc.; i factors com dies laborables on part de la població fa pont.
- L'algoritme s'estendrà en breu a tota la ciutat de Barcelona. Per això serà necessari disposar de les dades de treball de camp d'ocupació i de durada de l'estada de la resta de zones, no analitzades fins ara.
- Si el servidor de BSM no té problemes de capacitat de càlcul amb els models utilitzats aquí, es podrà explorar la conveniència d'incorporar models més elaborats, com el model AR(5), que té en compte l'ocupació dels 5 retards anteriors; o fins i tot, es podrà estudiar la possibilitat de deixar que el programa R trobi el model més escaient per si mateix (comanda `auto.arima`).
- Igualment, cas que el servidor tingui prou capacitat, es podrà avaluar quina és la periodicitat més adient per a cada predicció. Per exemple, per a una predicció a 30 minuts vista, podria ser més indicat un model que treballi amb periodicitat 30 minuts (que es faria servir per calcular la primera predicció que pot proporcionar), que un model de periodicitat 5 minuts (que tarda 6 períodes en calcular la predicció a 30



minuts). En tot cas, aquests canvis implicarien la necessitat d'estimar molts més models, i les proves fetes fins ara no demostren una millora massa significativa en les prediccions que s'obtenen.

- Quan la sèrie de registre sigui més llarga, es valorarà també la possibilitat de ponderar les observacions, de forma que les més recents tinguin més pes en el càlcul.
- Si el servidor de BSM no presenta problemes de capacitat i velocitat de càlcul, es valorarà la conveniència d'allargar la sèrie d'anàlisi, passant de 2 mesos a un període més llarg. S'ha comprovat que la longitud de la sèrie és un dels factors que més afecten al temps de procés necessari.
- Caldrà decidir què fer amb els dissabtes. Per aquests dies no hi ha treball de camp, i moltes de les sèries analitzades no tenen registres per als dissabtes, tot i ser dies en què la zona és operativa algunes hores. Incloure els dissabtes en les estimacions i prediccions amb aquestes condicions hauria distorsionat els resultats, i per això s'han omès.
- Les estimacions sobre ocupació no enregistrada, i sobre durada de l'estada, s'hauran d'anar revisant periòdicament. Especialment, el biaix entre l'ocupació registrada i l'ocupació total és molt possible que vagi disminuint a mida que els usuaris vagin entenent la necessitat d'efectuar correctament el registre.



Els continguts d'aquesta publicació estan subjectes a una llicència de **Reconeixement – No comercial – Sense Obra Derivada (by-nc-nd)** amb finalitat no comercial i sense obra derivada. Es permet copiar i redistribuir el material en qualsevol mitjà i format, sempre que no tingui finalitats comercials i no es distribueixi cap obra derivada.

La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.ca>