Escola Tècnica Superior d'Enginyers
de Camins, Canals i Ports de Barcelona
UNIVERSITAT POLITÈCNICA DE CATALUNYA

TESI DE MÀSTER

Màster

Enginyeria Civil (dins del programa de doctorat)

Títol

Desenvolupament i aplicació d'una metodologia per a la reconstrucció de camps de pluja diaris sobre tot Catalunya des de l'any 1941 fins al 2010

Autor

Joaquim Coll

Tutor

Daniel Sempere Torres

Intensificació

Data

Juliol del 2010

AGRAÏMENTS

En primer lloc vull agrair al meu tutor Daniel Sempere el suport en el desenvolupament del treball.

M'agradaria agrair també la tasca de tots els membres del Centre de Recerca Aplicada en Hidrometeorologia que dia a dia treballen per a que aquests tipus de projectes puguin tirar endavant.

També m'agradaria donar les gràcies a l'Agència Catalana de l'Aigua, el Servei Meteorològic de Catalunya i a l'Agència Estatal de Meteorologia per les dades subministrades.

Finalment, i de forma molt especial, a la Chus i l'Ariadna. El vostre suport fa que tot valgui la pena.

RESUM

L'obtenció de camps de pluja històrics amb una alta resolució espacial i temporal esdevé de vital importància per molts i diversos àmbits d'estudi. Resulta per tant rellevant establir les tècniques que ens permetin obtenir aquests camps de pluja de la forma més acurada possible.

El present treball té com a objectius l'establiment i aplicació d'una metodologia que permeti la reconstrucció de camps diaris de precipitació en el territori de Catalunya des de l'any 1941 fins el 2010.

Els mètodes tradicionals d'interpolació dels valors dels pluviòmetres permeten obtenir una primera aproximació dels valors del camp. Tot i així no permeten capturar de forma acurada la variabilitat espacial de la pluja que observem en l'actualitat quan utilitzem elements de percepció remota com son els radars meteorològics.

La metodologia presentada pretén fer ús d'aquesta informació per inferir-la en les observacions passades, i aconseguir camps de precipitació que tinguin en compte aquesta variabilitat espacial.

El primer pas de la metodologia consisteix en obtenir uns patrons de pluja fruït d'aplicar un Anàlisi de Components Principals (ACP) per a un període de calibració on tinguem uns bons camps de referència. En el nostre cas, els camps de precipitació utilitzats en com a referència són els obtinguts mitjançant la combinació de l'estimació del radar meteorològic amb la informació dels pluviòmetres mitjançant la tècnica de kriging amb deriva externa (KED), utilitzant com a període de calibració els dos anys en què es disposa d'aquest tipus de dades.

Els patrons obtinguts s'utilitzen com una base (els components de la qual són els vectors obtinguts fruït de l'ACP) en la qual representar els camps de pluja històrics. Els coeficients de la base per a cada dia a reconstruir s'extreuen de la informació que es té procedent dels registres històrics dels pluviòmetres, mitjançant una regressió multivariant en què les variables regressores són els valors de la base en els punts dels pluviòmetres i les observacions són el valor del pluviòmetre. Aquests coeficients ens serviran per expressar el valor del camp diari a estimar en la nova base de Components Principals.

Finalment, es combinen aquests camps estimats amb els registres puntuals dels pluviòmetres fent ús novament de la tècnica de kriging per tal de forçar que els camps finals reconstruïts tenen en el la situació en què es disposa d'un registre de pluviòmetre el valor d'aquest.

El mètode s'ha aplicat per a tota la sèrie de dades històriques des del 1941 i se n'han avaluat els resultats.

Si bé tècniques com la validació creuada mostren que quantitativament no s'obté una millora substancial sistemàtica respecte a l'estimació clàssica mitjançant

splines, si que qualitativament els camps reconstruïts semblen preservar la variabilitat espacial que es buscava. Això s'evidencia en comparar els diferents camps amb els obtinguts en episodis on es disposa d'una referència externa.

Paraules clau:

Pluja, interpolació, Anàlisi de Components Principals, kriging, regressió multivariant, radar meteorològic, pluviòmetre, Catalunya

ABSTRACT

Obtaining historical rainfall fields with high spatial and temporal resolution is vital to many different areas of study. It is therefore important to establish techniques that allow us to obtain these rainfall fields the most accurate possible.

The present work aims to establish and implement a methodology for the reconstruction of daily precipitation fields in Catalonia from 1941 until 2010.

Traditional methods of interpolation of the values from rain gauges provide a first approximation of field values. Still they cannot accurately capture the spatial variability of rain that we can observe nowadays when we use remote sensing instruments like weather radars

The methodology presented aims to use this information to infer it in past observations, and obtain rainfall fields that take into account this spatial variability.

The first step of the methodology is to obtain patterns of rain applying a Principal Component Analysis (PCA) for a calibration period where we have fields with a high quality and spatial resolution to use as a reference. In our case, the precipitation fields used as reference are obtained by combining the estimation of weather radar whit the measurements of the rain gauges through the technique of kriging with external drift (KED) performing the calibration period for the two years where we have this high quality data.

The patterns obtained are used as a basis (whose components are vectors obtained as a result of PCA) where we want to represent the historical rainfall fields. The coefficients of the basis for each day to reconstruct are extracted from the information that we have from the historical records of rain gauges. This is done through a multivariate regression, in which regressors are the values of the basin in the locations of the rain gauges and observations are the value of the rain gauge. These coefficients will then serve to express the value of the field to estimate in the new basis of Principal Components.

Finally, we combined these estimated fields with the specific records of rain gauges using the kriging technique again to force that this final reconstructed fields have in the locations where there is a register of rain gauge the value of this one.

The method was applied to the whole series of historical data from 1941 and the results have been evaluated.

While techniques like cross validation show that quantitatively there is not a substantial systematic improvement compared to the estimation using the classical methodology of splines, reconstructed fields using the new methodology seems to preserve the spatial variability that was searched. This is more evident when

comparing the fields with those obtained in current episodes where we have radar fields not used in calibration as external reference.

Keywords:
Rainfall, interpolation, Principal Components Analysis, kriging, multivariate regression, weather radar, rain gauge, Catalonia

ÍNDEX

1. Introducció 13

1.1. Objectius 13

1.2. Justificació 14

2. Metodologia 17

2.1. Tractament i disponibilitat de les dades 17

2.1.1. Xarxa de pluviòmetres 17

2.1.1.1 Dades disponibles 17

2.1.1.2 Distribució espacial i temporal de les dades pluviomètriques ... 19

2.1.1.3 Anàlisi de la qualitat de les dades pluviomètriques disponibles 20

2.1.2. Xarxa de radars 21

2.1.2.1 Dades disponibles 21

2.1.2.2 Anàlisi de les dades radar 21

2.1.2.3 Establiment dels millors camps de referència possibles 22

2.2. Obtenció de patrons de pluja: Anàlisi de Components Principals 22

2.3. Estimació dels camps de pluja històrics: Regressió Multivariant 28

2.3.1. Resolució mitjançant Mínims Quadrats 30

2.3.2. Singular Value Decomposition 31

2.4. Interpolació mitjançant kriging amb deriva externa (KED) 33

3. Resultats 35

3.1. Exemples de reconstrucció 35

3.2. Validació creuada 38

3.3. Comparació amb referència externa 42

4. Quantificació del temps emprat 49

5. Conclusions 51

5.1. Perspectives 52

6. Referències 55

1. Introducció

1.1. Objectius

L'objectiu d'aquest projecte és el desenvolupament i aplicació d'una metodologia que permeti la reconstrucció dels camps de pluja diaris sobre tot Catalunya des de l'any 1941 fins al 2010.

Tradicionalment els camps de pluja s'han obtingut mitjançant interpolació dels valors puntuals del pluviòmetres, ja sigui amb mètodes directes com les splines de làmina prima o amb tècniques geoestadístiques més avançades com ara el kriging. (els detalls d'aquestes tècniques estan disponibles en nombrosa literatura, com per exemple (Schumaker 2007) per a les splines, o (Deutsch; Journel 1997); (Isaaks; Srivastava 1989) per al kriging.). Els camps obtinguts d'aquesta forma tenen en els punts on hi ha mesura directe de pluviòmetre el valor d'aquesta, i a més proporcionen una estimació en tot el territori a través de minimitzar la curvatura del camp (en el cas de les splines) o per mitjà d'una interpolació lineal òptima (en el cas de tècniques com kriging). Si bé aquests camps són fàcils d'obtenir i suposen la referència tradicional, no són capaços de reproduir plenament la variabilitat de la precipitació observada amb instruments de percepció remota (com ara el radar meteorològic) que ofereixen una millor descripció del camp de pluja.

Aquest projecte proposa una metodologia que treu profit de les mesures d'alta resolució del camp de precipitació disponibles en l'època més actual (període de calibració) per tal d'extrapolar la variabilitat espacial al passat on aquestes mesures no estan disponibles. Així doncs, en l'època de calibració, mitjançant una Anàlisi de Components Principals (ACP) s'extreuen patrons de pluja característics amb alta resolució espacial, que són inferits al període històric de reconstrucció.

D'aquesta manera, sota la hipòtesi que l'estructura espacial de la precipitació té propietats invariants, l'aplicació de la metodologia permet la reconstrucció de camps de pluja a partir de la informació històrica disponible dels pluviòmetres.

La tècnica d'ACP fou inicialment desenvolupada per a l'anàlisi de dades atmosfèriques per Lorenz (Lorenz 1956) i està descrita en nombrosos llibres de referència (Wilks 2006).

D'altra banda, l'ús d'aquesta tècnica combinada amb d'altres per a la reconstrucció de sèries històriques de variables ha estat explorada per diferents autors. Així, hi ha diverses metodologies que es basen en la utilització de la informació d'una xarxa densa de dades disponibles en un determinat interval de temps per inferir aquesta informació en altres períodes de temps en que aquesta

xarxa és menys densa. En particular, la utilització de l'ACP i la posterior interpolació mitjançant diferents metodologies ha estat explorada en diferents estudis. Kaplan desenvolupà una tècnica (reduced space optimal interpolation (RSOI; (Kaplan et al. 1997)) en que s'utilitza la metodologia de les ACP per calibrar uns patrons i posteriorment realitzar el que es coneix com a interpolació òptima amb les dades d'aquells períodes en què no es disposa de prou informació. Aquesta tècnica fou després emprada per varis autors per a la reconstrucció històrica de diverses variables, com ara la reconstrucció de la precipitació acumulada mensual als Alps, (Ansell 2006) o la reconstrucció també a escala mensual de variables com la temperatura del mar (Rayner et al. 2003). Altres autors han utilitzat la mateixa metodologia per aplicar informació abundant de precipitació en el passat per a la millora de la detecció de la precipitació en el present (Schiemann et al. 2010).

La metodologia aquí descrita té en comú l'ús de l'ACP per a l'obtenció de patrons, però proposa una nova tècnica per a la reconstrucció. Així la metodologia desenvolupada obté inicialment uns camp de pluja estimats gràcies a combinar la informació històrica dels pluviòmetres amb els patrons obtinguts en el període de calibració. Després aquests camps estimats s'utilitzen com a deriva externa en la tècnica geoestadística anomenada Kriging amb Deriva Externa per fer la interpolació lineal òptima dels valors dels registres històrics dels pluviòmetres, obtenint d'aquesta manera uns camps reconstruïts històrics que tenen la textura dels camps de referència de tipus radar i el valor exacte dels pluviòmetres en aquells punts en què se'n disposa.

Per a la correcte implementació de la metodologia les dades de partença que s'han utilitzat han estat sotmeses a un anàlisi previ i acurat de validació i correcció d'errors. S'ha realitzat doncs un preprocessat tant de les dades radar (que s'utilizaran en el període de calibració per a l'obtenció de patrons característics de variabilitat espacial) com de les sèries pluviomètriques utilitzades per a la reconstrucció històrica dels camps.

1.2. Justificació

La reconstrucció de camps històrics de pluja esdevé de vital importància per molts i diversos àmbits d'estudi. Resulta per tant rellevant establir les tècniques que ens permetin obtenir aquests camps de pluja de la forma més acurada possible.

Es pretén per tant millorar les metodologies clàssiques que s'han utilitzat fins ara a Catalunya (interpolació mitjançant splines de làmina prima, tècniques geoestadístiques com el kriging) i aplicar-les sobre el període d'estudi.

Una de les principals aplicacions dels camps obtinguts rau de forma directe en l'ús que en fa la hidrometeorologia. Els camps obtinguts per tant permetran gràcies a la seva alta resolució espacial poder ser utilitzats per el càlcul de la precipitació acumulada en un cert interval de temps sobre una conca o subconca, de manera que

es puguin utilitzar els resultats per a la realització de sèries estadístiques de tot tipus, com ara la millora en la calibració de períodes de retorn de cabals dels rius de les Conques de Catalunya.

A més del seu ús en l'àmbit de la hidrometeorologia, els resultats obtinguts seran de gran servei evident per al seu ús en l'àmbit de la meteorologia pròpiament dita, i en general per a tots aquells àmbits d'estudi que precisin de camps de pluja històrics amb una alta resolució espacial i temporal.

A més, més enllà dels resultats, la metodologia exposada és vàlida per a ser emprada per a l'obtenció de resultats en qualsevol territori i espai de temps, sempre que les dades disponibles satisfacin els requeriments de la mateixa o en permetin l'adaptació.

2. Metodologia

S'ha desenvolupat una metodologia l'aplicació de la qual ens ha permès reconstruir camps de pluja acumulada diària des del 1941 fins a l'actualitat. Per a un correcte seguiment de la mateixa a continuació es resumeix de forma sintetitzada.

La metodologia té dues fases clares: la calibració i la reconstrucció.

En la fase de calibració es generen els millors camps de precipitació possibles durant els dos anys en els que es disposa de radars meteorològics, gràcies a la combinació d'aquests camps amb les dades de pluviòmetres. Un cop es tenen aquests camps de referència, s'extreuen uns patrons del comportament espacial de la pluja mitjançant un Anàlisi de Components Principals (veure apartat 2.2). Les dades radars de les que es disposa requereixen d'un preprocessat i validació, descrit en l'apartat 2.1.2.

En la primera fase de la reconstrucció els patrons obtinguts s'utilitzen com una base en la qual representar els camps de pluja històrics. Els coeficients d'aquesta base s'obtenen amb la informació dels pluviòmetres de cada dia. Això s'aconsegueix fent un a regressió multivariant, on les variables regressores són els valors de la base en els punts dels pluviòmetres i les variables dependents són els valors dels pluviòmetres. S'obtenen així els coeficients que serviran per a representar el camp a estimar en la nova base (veure apartat 2.3). En aquest punt s'utilitza la base de dades històrica de pluviometria amb resolució temporal diària de la que es disposa, la qual ha estat preprocessada per validar-ne les dades (veure apartat 2.1.1).

Com a últim pas per a la reconstrucció dels camps històrics s'utilitza la metodologia d'interpolació de kriging amb deriva externa per a combinar els camps estimats (usats com a deriva externa) amb el valor dels pluviòmetres, aconseguint així que els punts del camp on hi hagi pluviòmetre tingui exactament el valor d'aquest, i obtenint d'aquesta manera els camp reconstruïts finals (veure apartat 2.4).

2.1. Tractament i disponibilitat de les dades

2.1.1. Xarxa de pluviòmetres

2.1.1.1 Dades disponibles

Es disposa d'una base de dades amb registres de pluviòmetres des de l'any 1941 fins a l'actualitat. Aquests registres provenen de les quatre xarxes de pluviòmetres

que han coexistit en diferents períodes: la de l'Instituto Nacional de Meteorología (INM, actualment Agencia Estatal de Meteorología), la de l'Agència Catalana de l'Aigua (ACA), la de la Confederación Hidrogràfica del Ebro (CHE) i la del Servei Català de Meteorologia (SMC).

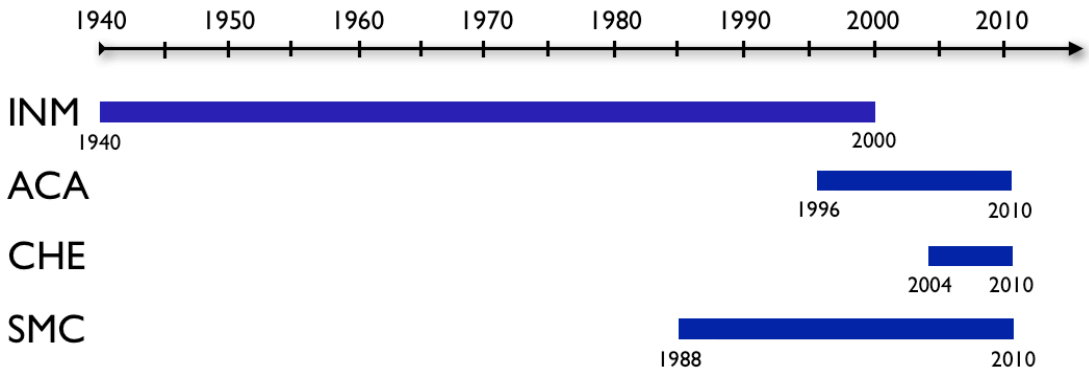

Figura 1. Mapa cronològic de les sèries pluviomètriques disponibles per l'estudi.

En la Figura 1 es presenta els anys en què es disposa de dades procedents de cada font. Inicialment només existia la xarxa de l'INM. Aquesta va desaparèixer l'any 2000 degut al trasllat de competències cap a les agències autonòmiques. Algunes estacions es van desmantellar i d'altres van passar a formar part de la xarxa del Sistema Automàtic d'Informació Hidrològica (SAIH), pertanyent a l'ACA. El SAIH comença el seu funcionament l'any 1996 i té com a funció monitoritzar les Conques Internes de Catalunya. Inclou en les seves estacions de control estacions meteorològiques, i en particular pluviòmetres.

Per la seva banda, el SMC (que també rep part dels pluviòmetres traspassats per l'INM l'any 2000) comença a operar l'any 1988, si bé no és fins a mitjans dels anys 90 quan aquesta xarxa augmenta la seva cobertura territorial i millora la seva gestió gràcies a les noves tècniques d'enregistrament i emmagatzematge automàtic .

Finalment, es disposa també de les observacions de la xarxa pertanyent a la Confederació Hidrogràfica de l'Ebre (CHE) a partir de l'any 2004.

Les diferents xarxes prenen registres de pluja en diferents resolucions particulars. Així, tenim per a cada xarxa la següent resolució:

1. INM: resolució diària.
2. SAIH-ACA: resolució 5-minutal
3. CHE: resolució 15-minutal
4. SMC: resolució 30 minutal o 60 minutal.

Les dades s'han homogeneïtzat per tal de poder treballar amb la resolució diària requerida en el present projecte.

2.1.1.2 Distribució espacial i temporal de les dades pluviomètriques

En un primer anàlisi de les dades s'ha estudiat la cobertura espacial dels pluviòmetres i el número d'aquests disponibles en cada any.

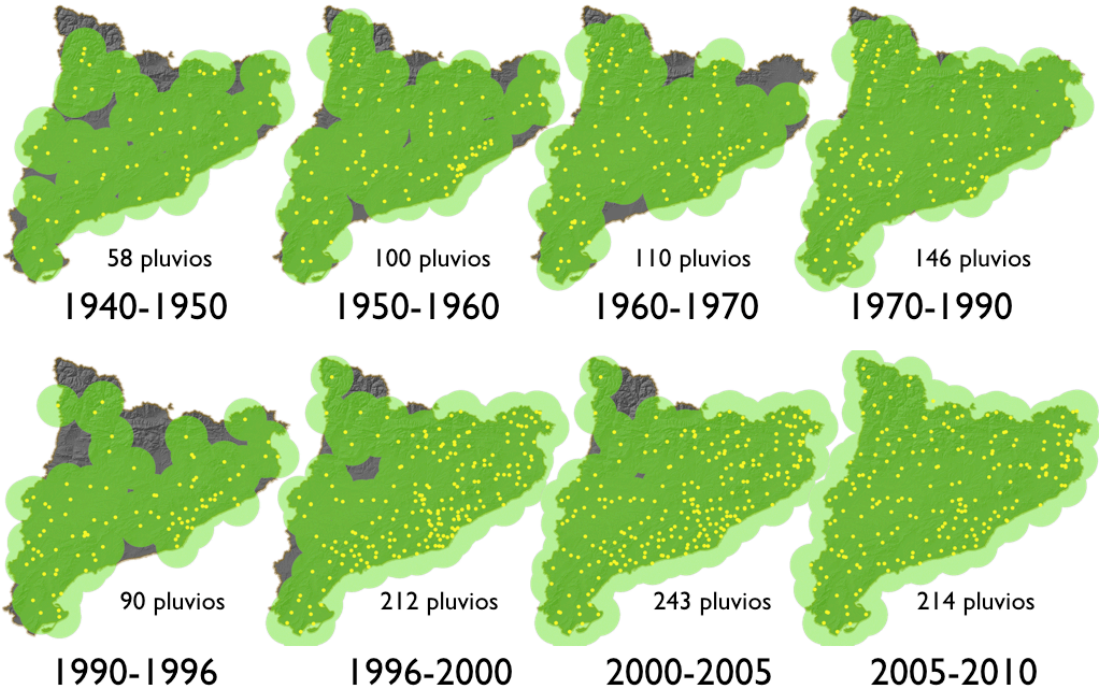

Figura 2. Distribució espacial dels pluviòmetres disponibles en el període de 1940 a 2010. Les àrees en verd indiquen les zones on hi ha, com a mínim, un pluviòmetre a menys de 20 km.

La Figura 2 mostra la mitjana de pluviòmetres disponibles agrupats en períodes i la seva distribució en el territori. La manera en què s'ha obtingut aquest resum és la següent: s'ha considerat un pluviòmetre com a vàlid si durant cadascun dels anys del període corresponent es disposava d'almenys 11 mesos de dades.

S'observa l'evolució de la densitat de les xarxes disponibles al llarga del temps. Aquesta ha anat augmentant al llarg dels anys, passant dels 58 pluviòmetres dels que es disposava en els anys 40 fins als 214 actuals. La distribució espacial dels pluviòmetres també ha millorat força. Això és especialment visible en passar del quinquenni corresponent al 200-2005 a la xarxa actual. Així, s'obté qua amb menys pluviòmetres actualment es disposa d'una molt bona cobertura del territori. Això ha estat així degut a la reestructuració que va fer l'ACA de la seva xarxa l'any 2005. Un altre punt a destacar és el pas del 1995 al 1996. Aquest any entra en funcionament la xarxa del SAIH de l'ACA, el que eleva molt considerablement el número de

pluviòmetres. La distribució d'aquests, molt concentrada en les Conques Internes, és degut a que el SAIH només opera en aquestes conques.

2.1.1.3 Anàlisi de la qualitat de les dades pluviomètriques disponibles

Les dades procedents dels pluviòmetres disponibles s'han sotmès a un procés de verificació i validació. Aquest procés ha estat diferent per a l'època de calibració, on es disposa dels camps radars coma referència, i per a l'època de reconstrucció, on no hi ha aquesta informació addicional.

En la calibració el procés de verificació ha estat especialment exhaustiu, doncs és molt important disposar dels millors camps de referència possible, i per això és necessari que les sèries pluviòmetriques en aquest període no presentin incoherències.

En el procés de calibració el que s'ha fet doncs és comparar cadascun dels camp obtinguts amb la metodologia de kriging de combinació de radar amb pluviòmetres (expicada en l'apartat 2.1.2) amb el camp corresponent a l'estimació mitjançant només radar. Aquest procediment ha permès detectar incoherències en les dades, com ara pluviòmetres amb registres de 0 mm quan el radar marcava pluja, registres de pluja elevada en un pluviòmetre en que el radar indicava absència de precipitació, o valors anormalment alts. Aquest procés ha permès identificar a nivell diari els pluviòmetres amb valors anòmals. El registre d'aquests pluviòmetres s'ha eliminat en el dia en què no han funcionat correctament.

La metodologia emprada per a la verificació de les dades de pluviometria històriques ha estat diferent, en no disposar d'una referència externa on comparar-se. Donat el volum de dades a verificar, els registres pluviomètrics diaris s'han acumulat per anys, i han estat aquestes acumulacions les que s'han analitzat. Aquestes acumulacions anuals s'han interpolat, el que ha permès realitzar un primer anàlisi visual de l'acumulació de cada any per a detectar incoherències flagrants, com ara absència total de pluja en un punt donat o acumulacions anormalment elevades. En una segona etapa, s'ha utilitzat el mètode del doble cúmul de masses (Searcy; Hardison 1960; Wilson 1990). Aquest mètode permet estudiar l'evolució de les sèries de dades d'un pluviòmetre determinat respecte les sèries de dades dels pluviòmetres més propers. D'aquesta manera és senzill determinar comportaments anòmals. Aquests registres incongruents s'han eliminat de la base de dades.

2.1.2. Xarxa de radars

2.1.2.1 Dades disponibles

Les dades radar utilitzades en el present projecte provenen de la xarxa de radars del Servei Meteorològic de Catalunya (SMC), composta actualment per 4 radars.

La xarxa de radars del SMC ha anat evolucionant des del primer radar instal·lat el 1996 a Vallirana (radar de Puig Bernat, PBE), fins a la instal·lació el l'any 2009 del quart radar, el de Miranda-Lleveria (LMI). Durant aquest període, s'han instal3lat el radar de Puig d'Arques(PDA, 2003) i el de Creu del Vent (CDV, 2004).

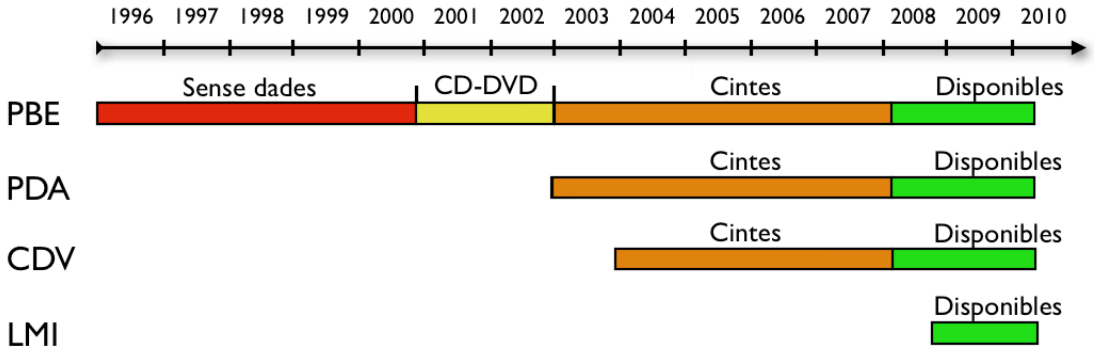

Figura 3. Evolució i disponibilitat de les dades dels radars del SMC.

La disponibilitat de les dades sense processar dels diferents radars es mostren en la Figura 3. Les dades tipificades com a disponibles son aquelles que estan emmagatzemades de forma que el seu accés és immediat, i son les que s'han pogut obtenir per a la realització del projecte. El període sens dades del radar de Puig d'Arques correspon a l'època en què s'utilitzava únicament amb finalitat científica i no se'n guardaven les dades. Per altra banda les dades en que l'emmagatzematge està en cintes o CD són de difícil accés i no s'han pogut utilitzar per al present estudi.

2.1.2.2 Anàlisi de les dades radar

Previ al seu ús, les dades dels diversos radars de la xarxa del SMC són processades mitjançant diferents metodologies per tal que l'estimació de la pluja sigui el més fidedigne possible. El processat de les dades inclou l'eliminació dels ecos de terra i els apantallaments orogràfics (Berenguer 2006; Sánchez-Diezma et al. 2001), l'extrapolació de la pluja al terra mitjançant un algorisme d'extrapolació per Perfils Verticals de Reflectivitat (Franco 2008), i la transformació dels valors de reflectivitat en pluja a través d'una relació Z-R (Sempere-Torres et al. 1994). L'acumulació dels camps puntuals de pluja en període escales temporals més grans com la diària es realitzat tenint en compte el moviment dels camps de precipitació a través de tècniques de morphing. A més, es calibren els resultats amb una referència

externa, en aquest cas la xarxa de pluviòmetres (Rosenfeld et al. 1994; Wilson; Brandes 1979)

Aquest processat es realitza doncs per als quatre radars disponibles. Malauradament, en molts casos la calibració final indica que la mesura del radar està excessivament allunyada de la referència. Això és degut a diversos factors coneguts ocorreguts radars (PDA, PBE) no se'n pugui utilitzar cap dada. L'anàlisi dels radars restants (CDV, LMI) estableix un període de dades fiables de dos anys, el comprès entre Març del 2008 a Febrer del 2010, i que serà l'utilitzat com a període de calibració.

2.1.2.3 Establiment dels millors camps de referència possibles

Disposem per tant de dues fonts de dades de tipus radar ja preprocessades. Per a obtenir el millor camp possible el que fem és una composició de radars i posteriorment una combinació del resultat amb les dades dels pluviòmetres.

La composició dels dos radars en una sola imatge s'obté considerant el valor màxim dels dos radars allà on aquests es solapen, i el valor corresponent a cadascun en la resta del domini.

Per la seva banda, el camp radar obtingut es combina amb els pluviòmetres disponibles mitjançant la metodologia d'interpolació de kriging amb deriva externa (Velasco-Forero et al. 2009). Aquesta metodologia utilitza el camp radar com a deriva externa per a fer una interpolació per kriging entre els valors dels pluviòmetres. En ser la metodologia del kriging un interpolador exacte, imposem la condició que el camp resultant tingui en els punts en què hi ha registre de pluviòmetre el valor d'aquests.

D'aquesta manera obtenim uns camps de pluja diària que son els millors que es poden obtenir fent ús dels instruments de mesura de precipitació actuals (pluviòmetres i radars meteorològics) durant els dos anys dels que disposem dades amb suficient qualitat. Aquests dos anys per tant constituiran el nostre període de calibració.

2.2. Obtenció de patrons de pluja: Anàlisi de Components Principals

La metodologia per a la reconstrucció de camps de precipitació històrics té com a primer pas la identificació dels patrons espacials de precipitació en una època en la que es disposi de mesures d'alta resolució.

Com s'ha vist en aquest treball s'ha utilitzat com a període de calibració els dos anys entre l'1 de Març de 2008 i el 28 de Febrer de 2010, i com a producte d'alta

resolució s'ha utilitzat la combinació de les estimacions de la xarxa de radars meteorològics del SMC amb les estimacions puntuals de pluviòmetres mitjançant la metodologia de kriging amb deriva externa comentada anteriorment. S'ha treballat a una resolució espacial de 3 x 3 km^2 com a compromís entre mantenir una resolució adequada i la correcte resolució dels problemes numèrics que comporta treballar amb matrius de dades tant grans. El pas de la resolució d'1 x 1 km^2 de les dades originals de radar combinades amb pluviòmetres a la resolució de treball s'ha fet promitjant els valors dels 9 píxels.

Per a la obtenció dels patrons espacials de pluja, s'ha realitzat un Anàlisi de Components Principals (ACP) sobre els dos anys de dades.

L'ACP consisteix en descomposar les observacions (camps de pluja diaris) en una nova base ortonormal amb unes característiques determinades que la fan avantatjosa per al nostre propòsit. El primer avantatge que té aquesta descomposició és que els vectors d'aquesta nova base estan ordenats de manera que cadascun explica de major a menor variància del camp original. L'altre avantatge fonamental és que en estar aquesta base construïda en base a les pròpies observacions aquests vectors més representatius aconseguiran respondre a patrons que són els que ens interessa reproduir.

L'espai de Components Principals s'obté d'extreure els Vectors Propis de la matriu de covariància de les observacions centrades i reduïdes. D'aquesta manera, si definim $p_{n,m}$ com el valor corresponent a l'observació del dia n en el píxel m (d'un total de N dies en el període de calibració) podem definir la nostra variable centrada i reduïda com a:

$$x_{n,m} = \frac{p_{n,m} - \mu_m}{\sigma_m} \qquad \text{(Eq. 1)}$$

on:

$x_{n,m}$: Valor estandarditzat del dia n en el píxel m

$p_{n,m}$: Valor del dia n en el píxel m

μ_m: mitjana del píxel m en tot el període de calibració

σ_m: desviació estàndard del píxel m en tot el període de calibració

Per tant la nostra matriu de covariancia la calcularem de la forma:

$$S_x = (N-1)^{-1} \sum_{n=1}^{N} x_n x_n^T \qquad \text{(Eq. 2)}$$

on:

N: Número de d'observacions

x_n: Vector columna de tots els valors dels píxels estandarditzat x en el dia n

En el nostre cas, en tenir dos anys de dades en el període de calibració, el nombre inicial d'observacions N és de 731. De tota manera, s'ha decidit eliminar aquells dies amb molt poca o gens de pluja. Incloure o no els dies sense pluja no varia el resultat del càlcul dels vectors propis, però en canvi incrementa temps de càlcul. Els dies amb molt poca pluja, per a la seva banda, no aporten informació als patrons de pluja i també incrementen el cost final computacional.

Establint doncs un balanç entre el temps de càlcul necessari i la informació realment transcendent s'han utilitzat per a l'estudi aquells dies en què s'ha observat almenys 1 mm de pluja acumulada en més d'1% del territori, obtenint finalment un número d'observacions N=439

El nombre total de Components Principals (que anomenarem M) és per força igual al nombre de píxels que hi ha en cada camp. L'estudi s'ha realitzat per al territori de Catalunya amb una resolució de 3 x 3 km2 i per tant tenim els camps tenen un número de píxels M=3427.

Podem per tant expressar cada observació de la següent manera:

$$Observació_n = \sum_{j=1}^{j=M} \alpha_{n,j} CP_j \qquad \text{(Eq. 3)}$$

on:

$\alpha_{n,j}$: Coeficient de la CP *j* per a la observació corresponent al dia n

CP_j: Component Principal j

M: Número total de CP

I de forma matricial:

$$\underset{(Mx1)}{x_n} = \underset{(MxM)}{E} \underset{(Mx1)}{a_n} \qquad \text{(Eq. 4)}$$

on:

x_n: Camp observat

a_n: Matriu de coeficients

E: Matriu de Components Principals

Cal dir que tant en l'equació 3 com en la 4 el camp resultant en el nostre cas s'ha obtingut després de centrar i reduir la variable original (veure equació 1) de manera que caldrà a posteriori la seva transformació per obtenir els camps en la seva forma real (veure equació 18).

Hem vist que les CP són directament els Vectors Propis de la matriu de covariancia S_x, mentre que els coeficients de cada component són la projecció de l'observació en la nova base de CP. Per tant en la calibració en tant que coneixem les observacions els coeficients de cada component es poden calcular directament de la següent forma:

$$\underset{(Mx1)}{a_n} = \underset{(MxM)}{E^{-1}} \underset{(Mx1)}{x_n} \qquad \text{(Eq. 5)}$$

Com s'ha comentat, un dels avantatges d'utilitzar components principals com a base és que agafant un número molt reduït de vectors som capaços d'explicar gran camp de la variància inicial del camp.

La Figura 4 mostra la variància explicada per cada CP de la nostra base, i la seva funció d'acumulació. Es veu com els CP que s'obtenen de l'anàlisi estan ordenats de manera que cadascun explica en mitjana menor mitjana que l'anterior. En el gràfic amb l'acumulació s'observa que per explicar de mitjana un 95% de la variància del camp observat n'hi ha prou amb agafar les 78 primers CP, mentre que si n'agafem 148 ja estem explicant el 99% de la variància. En ambdós casos hem reduït la nostra base a un nombre molt menor d'elements, doncs inicialment disposavem d'un espai de 3427 vectors. Això serà d'especial importància en el procés de reconstrucció dels camps, explicats en l'apartat 2.3.

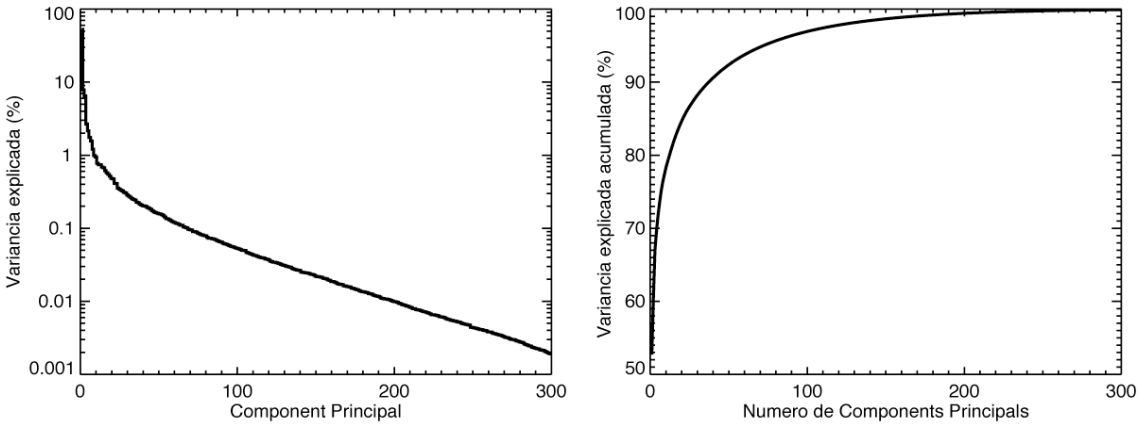

Figura 4: La figura de l'esquerra mostra la variància explicada per cadascuna de les 300 primeres CP obtingudes per al període de calibració. La figura de la dreta és l'acumulació del mateix valor.

Ara per tant podem obtenir uns nous camps de pluja que anomenarem camps filtrats, que representen les observacions en funció del nombre de CP que agafem. Així, l'equació 1 queda ara com a:

$$Observació_Filtrada_n = \sum^{j=L} \alpha_{n,j} CP_j \qquad \text{(Eq. 6)}$$

on L és el número de CP que hem escollit.

I de forma matricial:

$$\underset{(Mx1)}{x_n^f} = \underset{(MxL)}{E'} \underset{(Lx1)}{a_n'} \qquad \text{(Eq. 7)}$$

on:

x_n^f : Camp del dia n filtrat per a L Components Principals

a_n': Matriu dels coeficients corresponents a les L primeres CP

E': Matriu de les L primeres Components Principals

Altra cop, caldrà desfer la transformació inicial per obtenir els camps de pluja (multipliquem per la desviació estàndard i afegim la mitjana que hem tret en l'equació 1).

La Figura 5 mostra el camp de pluja del dia 10 de maig de 2008 reconstruït agafant un número determinat de CP. S'observa que a mesura que incrementem L, el camp filtrat és cada cop més semblant a l'original. En el cas de L=78, el camp visualment és pràcticament idèntic a l'observació (de fet podem definir l'observació com a un cas particular del camp filtrat en que L=M=3427).

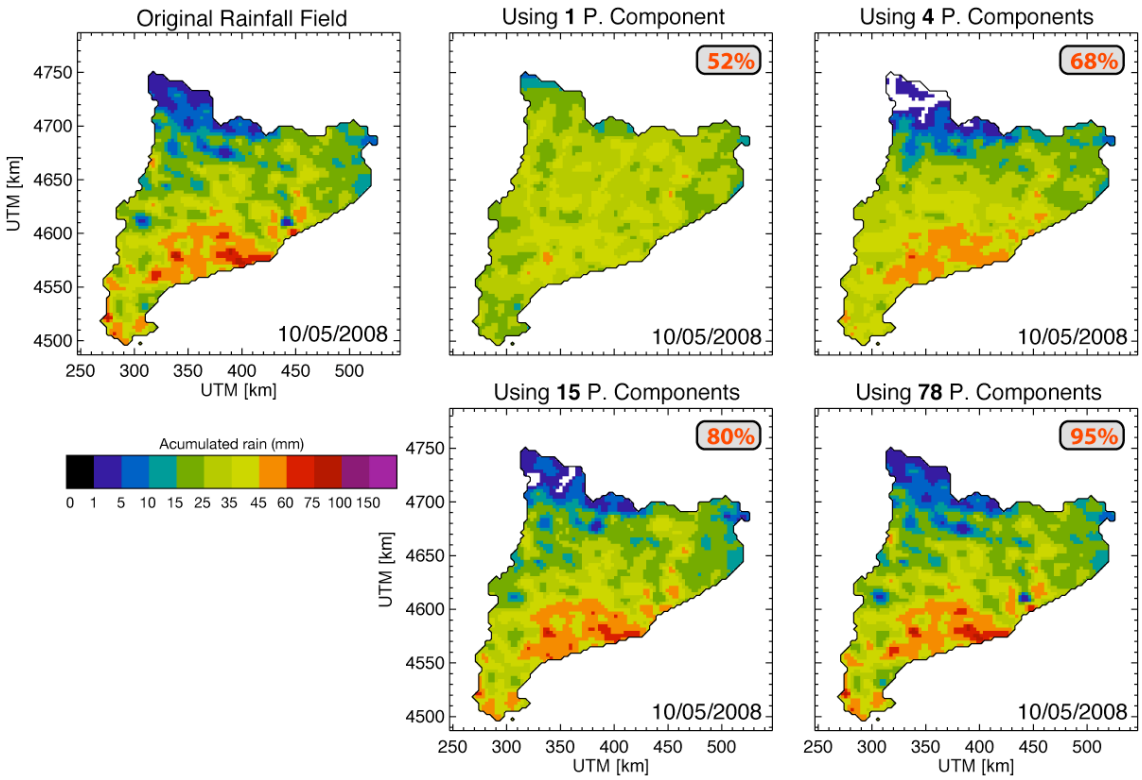

Figura 5: Efecte del filtratge del camp original prenent diferent número de CP. A dalt a l'esquerra hi ha el camp de precipitació observat el dia 10 de Maig de 2008 obtingut mitjançant les mesures de radar i pluviòmetres. Quatre

representacions del mateix camp de precipitació en la base de components principals prenent 1, 4, 15 i 78 components principals. En cada imatge està escrit el percentatge de la variància explicada en promig.

Com hem comentat la segona característica que fa especialment indicat l'ús de l'Anàlisi de components Principals és que els components de la base, en estar feta a partir de les pròpies dades, defineixen patrons de comportament d'aquestes dades. Això és especialment evident en les primeres CP, aquelles que representen més variància.

La Figura 6 mostra les quatre primeres CP. En ella es veuen clarament uns patrons de comportament: en la primera CP, que representa per ella sola un 52 % de la variància dels camps, es pot observar la variabilitat a petita escala de la pluja. La segona CP mostra el gradient Nord-Sud , mentre que la tercera CP mostra la variabilitat llevant-ponent (costa-interior). Ambdues CP representen respectivament el 7 i el 6 % de la variància. La quarta CP representa ja només un 2% de la variància, de manera que és més complicat trobar una raó evident per al patró presentat. Podria tractar-se de la variabilitat entre la Catalunya central i la resta, però també l'efecte d'atenuació del radar amb la distància que la combinació amb els pluviòmetres no hauria pogut acabar de mitigar.

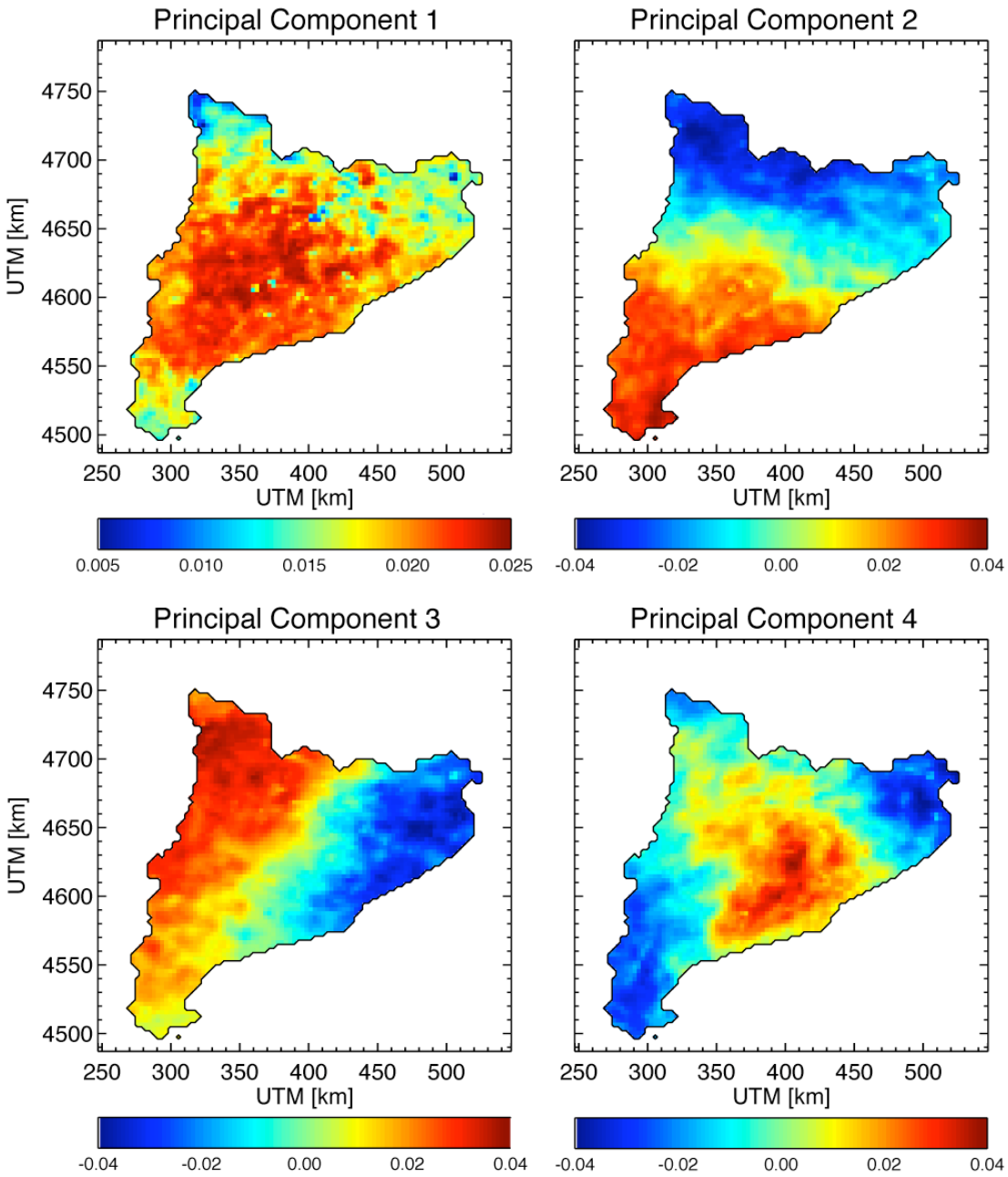

Figura 6: Representació gràfica de les 4 primeres CP corresponents als dos anys de calibració. Unitats arbitràries.

2.3. Estimació dels camps de pluja històrics: Regressió Multivariant

Per a la reconstrucció dels camps de precipitació històrics del 1941 al 2007 el nostre objectiu serà representar els camps en la base de CP calculada en el període de calibració, utilitzant la informació dels pluviòmetres corresponent a cada dia de reconstrucció per calcular el pes (o coeficient) que ha de tenir cada element de la nostra base.

Cal remarcar que la nostra base sempre és la mateixa per a tots els camp reconstruïts, i que és aquesta la que ens aporta la informació sobre els patrons de pluja capturats en el període de calibració. Prenem doncs com a hipòtesi de partença que els patrons de pluja són invariants al llarg dels anys.

Podem dir per tant que el que volem obtenir és el següent:

$$Observació_Estimada_{n'} = \sum_{j=1}^{j=L} \alpha_{n',j} CP_j \qquad \text{(Eq. 8)}$$

On ara tenim com a incògnites tant l'observació del camp el dia n' com els coeficients $\alpha_{n'}$

De forma matricial podem escriure:

$$\underset{(Mx1)}{x_{n'}^{f}} = \underset{(MxL)}{E'} \underset{(Lx1)}{a'_{n'}} \qquad \text{(Eq. 9)}$$

on:

$x_{n'}^{f}$: Estimació del camp el dia n', filtrat per L Components Principals

E': Matriu de les L primeres Components Principals

$a_{n'}'$: Vector de coeficients a obtenir

En tant que tenim la informació exacte de la pluja en uns punts concrets del camp a estimar, és a dir, la informació dels pluviòmetres, podem estimar les nostres $\alpha_{n'}$ en aquests punts concrets i utilitzar posteriorment aquests coeficients per a tot el camp a reconstruir.

Definim en l'equació 10 que el valor de cada pluviòmetre (tractat com es veurà seguidament) es pot expressar com una combinació lineal del valor de la base reduïda dels CP en el punt del pluviòmetre més un cert error.

Si definim que aquesta relació entre els pluviòmetres i els diferents valors de les CP ve definida per uns coeficients que considerem constants per a cada camp, tindrem un sistema d'equacions sobreestimat on la variable independent (o regressora) és el valor de les CP en la localització dels pluviòmetres, la variable depenent (o observació) és el valor dels pluviòmetres amb un tractament inicial en el dia a considerar, mentre que les incògnites són els coeficients a estimar:

$$G_{k,n'} = \sum_{j=1}^{j=L} \hat{\alpha}_{n',j} CP_j(x_k) + \varepsilon_{n',k} \qquad \text{(Eq. 10)}$$

on:

$G_{k,n'}$: Valor tractat del pluviòmetre k en el dia n'

$\hat{\alpha}_{n',j}$: Estimació del coeficient de la CP j per a la observació corresponent al dia n'

$CP_j(x_k)$: Valor de la CP j en la localització del pluviòmetre k (x_k)

$\varepsilon_{n',k}$: Error en l'estimació per al pluviòmetre k en el dia n'

Com hem comentat inicialment tractem la variable depenent. El que fem és restar al valor del pluviòmetre una mitjana i dividir per una desviació estàndard. La mitjana i la desviació estàndard que utilitzem és la que hem calculat de forma prèvia en la calibració per a dos anys de dades i per a tots els punts del territori., que és la mateixa que s'ha utilitzat per transformar les dades inicialment abans de calcular les CP (veure equació 1). El valor utilitzat en cada pluviòmetre serà el corresponent a la seva situació en el territori.

D'aquesta manera tenim que:

$$G_{k,n'} = \frac{(p_{k,n'} - \mu(x_k))}{\sigma(x_k)} \qquad \text{(Eq. 11)}$$

on;

$p_{k,n'}$: Valor del pluviòmetre k en el dia n'

$G_{k,n'}$: Variable tractada del pluviòmetre k el dia n'

$\mu(x_k)$ Valor de la mitjana μ en la localització del pluviòmetre k

$\sigma(x_k)$: Valor de la desviació estàndard σ en la localització del pluviòmetre k

El que fem és dimensionar el sistema presentat en l'equació 10 per tal de tenir menys incògnites que equacions i poder així resoldre'l d'una forma acurada; d'aquí la importància cabdal de poder reduir l'espai de CP original a un subespai molt més petit del qual en coneixem la variància que estem representant.

Després de vàries anàlisis amb diferent número de CP i de pluviòmetres, s'ha obtingut com a número de CP que funciona de forma correcte per a tots els casos el número 30. Aquest número de CP captura bé l'estructura dels camp ja que representa de mitjana el 90% de la variància dels camps originals, i permet que fins i tot en el cas de tenir pocs pluviòmetres com en les dècades dels 40 o 50, la regressió obtingui uns resultats acurats.

2.3.1. Resolució mitjançant Mínims Quadrats

Hi ha diversos mètodes de regressió per a resoldre un sistema d'equacions sobredimensionat.

Sigui doncs el sistema presentat en l'equació 10. La distribució de la variable aleatòria G és per hipòtesi normal, de manera que el model de màxima verosimilitud equival a mínims quadrats, és a dir, es demostra que el problema es resol de forma òptima si es té en compte minimitzar l'error mitja quadràtic (MSE).

Si describim el sistema d'equacions en forma matricial tenim:

$$\underset{(NGx1)}{P_{n'}} = \underset{(NGxL)}{E_i}\ \underset{(Lx1)}{\hat{a}_{n'}} + \underset{(NGx1)}{U_{n'}} \qquad \text{(Eq. 12)}$$

on:

NG: Número de pluviòmetres el dia n'

L: Número de CP escollides

$P_{n'}$: Vector amb els valors tractats dels pluviòmetres el dia n' (variable depenent o resposta)

E_i : Matriu dels valors de les L components principals en les localitzacions dels pluviòmetres (variable independent o regressora)

$\hat{a}_{n'}$: Vector de coeficients de les L components principals per al dia n'

$U_{n'}$: Vector dels residus per al dia n'

Es demostra que la resolució dels estimadors pel mètode de mínims quadrats és de la forma:

$$\underset{(Lx1)}{\hat{a}_{n'}} = (\underset{(LxNG)}{E_i^T}\ \underset{(NGxL)}{E_i})^{-1}\ \underset{(LxNG)}{E_i^T}\ \underset{(NGx1)}{P_{n'}} \qquad \text{(Eq. 13)}$$

on $\hat{a}_{n'}$ es diu que és un BLUE, de les sigles en anglès Best Linear Unbiased Estimate (millor estimador lineal no esbiaixat)

2.3.2. Singular Value Decomposition

El principal inconvenient d'utilitzar un resolució del sistema mitjançant el procediment de l'equació 12 es presenta quan si tenim problemes de colinealitat entre les variables regressores, i com veurem, aquest és el cas que ens ocupa.

Es defineix un conjunt de variables com a colineals si una d'elles és combinació lineal de les altres. Són quasi-colineals quan una d'elles és aproximadament una combinació lineal de les altres. En general utilitzarem el terme colineals per a referir-nos a ambdós casos.

Així doncs dues variables són colineals si estan altament correlacionades. En les dades d'aquest estudi ens trobem que sovint les diverses variables independents estan correlacionades entre elles. Això fa que en utilitzar la metodologia d'estimació per mínims quadrats els resultats siguin poc acurats.

L'explicació matemàtica del problema rau en el fet que el producte matricial $E_i^T E_i$ de l'equació 13 esdevé quasi singular, de manera que la seva inversió esdevé en resultats poc acurats.

Una manera de controlar millor el resultat de l'estimador $\hat{a}$ és utilitzant la Singular Value Decomposition (SVD), la qual consisteix a trobar un mètode de manera que puguem calcular de forma directe els estimadors mitjançant:

$$\underset{(Lx1)}{\hat{a}_{n'}} = \underset{(LxNG)}{E_i^{-1}} \underset{(NGx1)}{P_{n'}} \qquad \text{(Eq. 14)}$$

Com que E_i no és quadrada no es pot invertir. La SVD s'utilitza per a trobar el que s'anomena una pseudo-inversa de E de manera que siguem capaços de resoldre l'equació 14.

La SVD ens diu que tota matriu es pot descomposar de la forma:

$$\underset{(LxNG)}{E_i} = \underset{(LxL)}{U} \underset{(NGxNG)}{W} \underset{(NGxNG)}{V^T} \qquad \text{(Eq. 15)}$$

on:

U: Matriu amb els vectors singulars del costat esquerra

W: Matriu diagonal amb els valors singulars en la diagonal

V: Matriu amb els vectors singulars del costat dret

La característica especial d'aquesta descomposició és que W conté els valors singulars de la matriu E ordenats de major a menor variància explicada.

D'aquesta manera la matriu W es pot disminuint de rang prenent només els valors singulars amb més variància, i deixant fora aquells amb menor variància.

De forma semblant al que fa l'Anàlisi de Components Principals, ens podem quedar doncs amb aquells valors singulars que representen més variància (i els seus vectors propis associats).

Una vegada hem descomposat la nostra matriu de la forma de l'equació 15, podem ara calcular-ne la pseudo-inversa, prenent tots els valors de W o com s'ha comentat reduïnt-ne el rang i prenent només aquells que representen major variància. La pseudo-inversa tindrà per tant la forma:

$$\underset{(NGxL)}{E_i^{-1}} = \underset{(NGxNG)}{V} \underset{(NGxNG)}{W^{-1}} \underset{(LxL)}{U^T} \qquad \text{(Eq. 16)}$$

De manera que ja podrem obtenir els estimadors $\hat{a}$ mitjançant l'equació 13.

Anant de nou a l'equació 8 podem ara substituir els valor incògnita α per aquestes estimacions fruit de la resolució de la regressió múltiple, $\hat{\alpha}$, tenint en compte però que hem de desfer la transformació feta a l'equació 1.

Per tant sabem que el nostre camp estimat el podem obtenir resolent la següent equació:

$$\frac{(Observació_Estimada_{n'} - \mu)}{\sigma} = \sum_{j=1}^{j=L} \hat{\alpha}_{n',j} CP_j \qquad \text{(Eq. 17)}$$

on:

μ: Vector que conté el valor de la mitjana en cada punt del camp calculada en la calibració

σ: Vector que conté el valor de la desviació estàndard en cada punt del camp calculada en la calibració

I tindrem per tant finalment:

$$Observació_Estimada_{n'} = \sum_{m=1}^{m=M} \sum_{j=1}^{j=L} \frac{\hat{\alpha}_{n',j} CP_{j,m} + \mu_m}{\sigma_m} \qquad \text{(Eq. 18)}$$

on:

μ_m: Valor de la mitjana en el punt m calculada en la calibració

σ_m : Valor de la desviació estàndard en el punt m calculada en la calibració

$CP_{j,m}$: Valor de la CP j en el punt m

$\hat{\alpha}_{n',j}$: Estimador del coeficient de la CP j en el dia n'

L: Número de CP

M: Número total de píxels en el camp

2.4. Interpolació mitjançant kriging amb deriva externa (KED)

Els camps obtinguts mitjançant l'equació 17 ens donen ja una bona estimació del camp. Tot i així, s'ha volgut imposar la condició que en els punts en què hi ha

pluviòmetre, el valor del camp estimat en aquell punt sigui exactament el valor registrat per el pluviòmetre, la qual cosa el sistema proposat fins ara no satisfeia.

La metodologia utilitzada és anàleg a la presentada en el punt 2.1.2. En aquest cas, s'han utilitzat els camps ja estimats com si fossin observacions radars, i s'han combinat amb els pluviòmetres mitjançant l'interpolador exacte que és el mètode de kriging amb deriva externa (KED).

En aquest últim pas de la metodologia, doncs, s'ha aconseguit imposar com a deriva externa de la metodologia del KED els camps estimats, i obtenir uns nous camps (que anomenarem reconstruïts, per diferenciar-los semànticament) que satisfan la condició de tenir en els punts on es disposa de pluviòmetres el valor d'aquests.

Els resultats presentats en l'apartat 3 i la validació de la metodologia mostren com aquest darrer pas millora significativament els camps reconstruïts.

3. Resultats

S'ha utilitzat la metodologia descrita per a reconstruir els camps de pluja diària de 68 anys que comprèn l'estudi. Es disposa per tant de 24.721 camps de pluja.

En el present document es presenten casos específics dels diferents camps reconstruïts i s'analitzen els resultats obtinguts d'aplicar diferents metodologies d'avaluació, verificació i contrast.

3.1. Exemples de reconstrucció

La Figura 8 mostra l'exemple de reconstrucció del camp de pluja per al dia 18 de gener de 1962. El camp reconstruït final es compara amb el que s'obtindria d'obtenir una interpolació directe dels pluviòmetres mitjançant splines de làmina prima, i amb el que s'obtindria també si no s'apliqués la combinació final amb pluviòmetres mitjançant la tècnica de kriging explicada en l'últim pas de la metodologia.

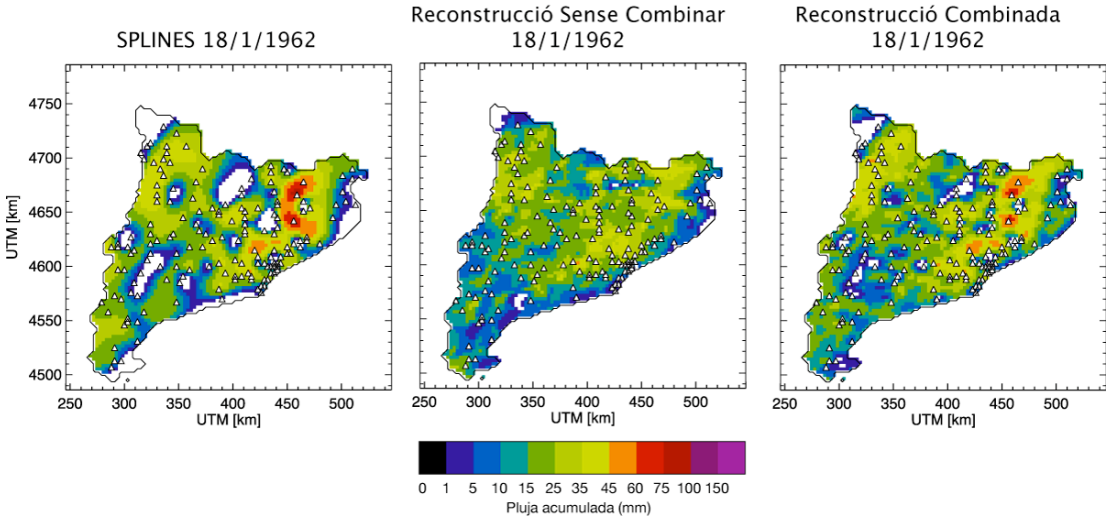

Figura 7: Camp reconstruït per al dia 18 de Gener de 1962. La primera figura de l'esquerra correspon a la interpolació directe amb splines. La segona figura representa el camp estimat sense combinar amb pluviòmetres. La tercera figura és el camp reconstruït final un cop s'ha combinat amb els pluviòmetres utilitzant la tècnica de kriging amb deriva externa

L'exemple reconstruït correspon a un dia plujós però no extrem, en que es disposava de 162 pluviòmetres. Es mostra en la figura com el camp reconstruït amb la tècnica presentada en el present projecte aconsegueix preservar la textura característica dels camps obtinguts mitjançant l'ús dels radars meteorològics. La

interpolació clàssica mitjançant splines, per contra, mostra les formes arrodonides característiques d'aquesta metodologia. Observem també que la interpolació final amb kriging aconsegueix matissar les zones en què els pluviòmetres indicaren que no va ploure tant , i sobretot que emfatitzar aquelles zones on si que ho va fer. Això es veu clarament en la zona de la Garrotxa on el camp combinat reprodueix els valors de pluja intensa (al voltant de 100 mm) que no havíem aconseguit reproduir amb el camp sense combinar.

La Figura 8 mostra l'exemple de reconstrucció per un cas de pluja extrema. És l'event del 7 de novembre de 1982, que provocà força inundacions al Pirineu, especialment greus al poble de Pont de Bar que va ser destruït completament.

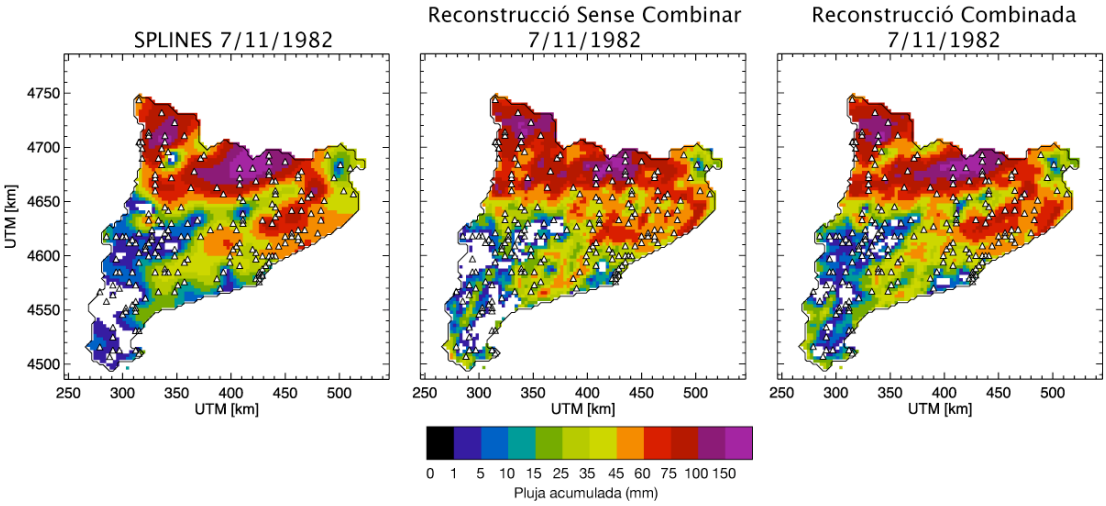

Figura 8: Ídem Figura 7 però per al dia 7 de Novembre de 1982.

L'exemple d'aquesta figura mostra uns camps molt semblants entre ells. En aquest cas es disposava de 193 pluviòmetres, el que significa una xarxa densa de pluviometria, que afavoreix tant al mètode d'splines com a la metodologia presentada. De tota manera la reconstrucció mitjançant la tècnica exposada reprodueix millor la textura d'un camp de precipitació. Les splines segueixen mostrant formes més arrodonides.

També s'aprecia que la combinació final aconsegueix ajustar millor el camp reconstruït sobretot en aquelles regions on els pluviòmetres han registrat pluja. Això és especialment important en la zona dels Pirineus, on es reprodueixen els valors extrems de pluviòmetres que va arribar a enregistrar més de 250 mm acumulats en aquell sol dia. Aquest efecte també es pot apreciar en les comarques del Segrià i del Baix Camp, on els pluviòmetres van indicar pluja (veure splines), el nostre camp estimat no en va considerar, però el camp reconstruït final en tenir en compte el valor dels pluviòmetres en aquelles zones si que hi ha reproduït pluja.

El tercer exemple mostrat correspon al d'un dia poc plujós, el 24 de març de 1962 (Figura 9) en què es disposa de 160 pluviòmetres.

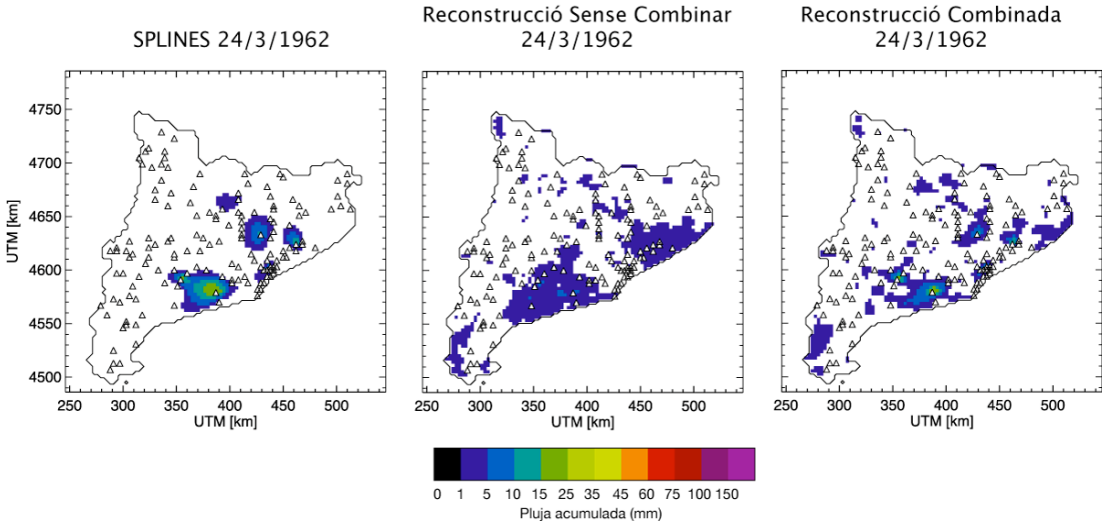

Figura 9: Ídem Figura 7 però per al dia 24 de Març de 1962.

Els dies pocs plujosos o amb pluja espúria molt localitzada és on les diferents metodologies esdevenen en resultats menys satisfactoris. En aquest cas, per exemple, només sis pluviòmetres han registrat pluja. Els splines l'han localitzat tota en una àrea molt petita corresponent a l'àrea d'influència de cada pluviòmetre. La metodologia sense combinar, seguint els patrons que li hem inferit, ha extès aquesta pluja per més territori de forma potser més creïble, però a la vegada indica pluja en llocs allunyats dels focus de precipitació que no semblen reals (Vall d'Aran, Baix Ebre). Val a dir però que aquesta pluja és molt i molt minse (de l'ordre d'1mm) de manera que si bé visualment no és, qualitativament té molt poca importància. La combinació final, per la seva banda, ha intensificat les zones en que hi ha pluviòmetres amb el valor d'aquests, i ha eliminat la pluja mina allà on hi ha pluviòmetre i aquest ha marcat zero (per exemple en el Tarragonès).

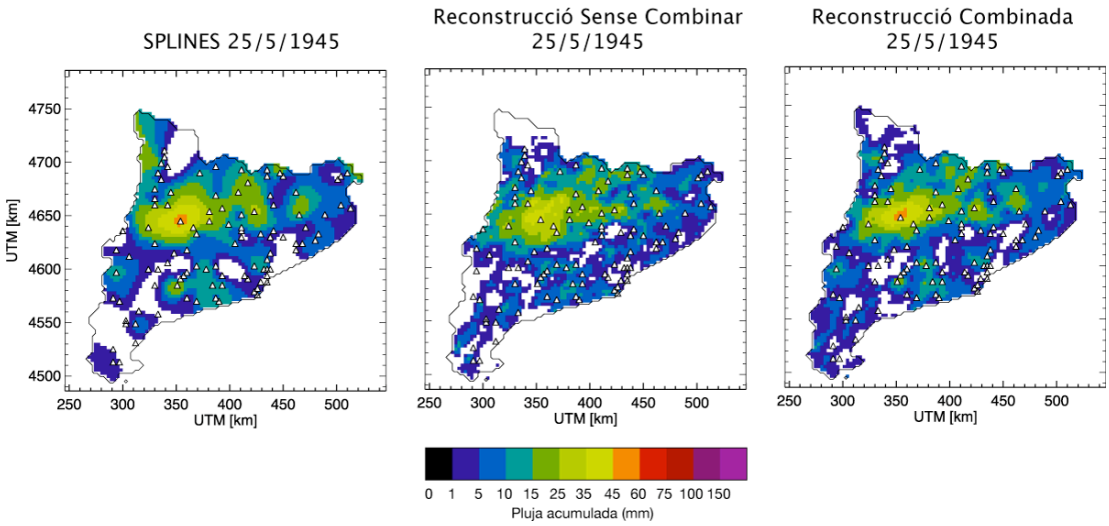

Figura 10: Ídem Figura 7 però per al dia 25 de Maig de 1945.

Finalment, es mostra en la Figura 10 un exemple de reconstrucció per al dia 25 de Maig de 1945, un dia pertanyent a una dècada en què la xarxa de pluviòmetres era molt menys densa, amb uns 100 pluviòmetres de mitjana. Concretament per aquest cas es disposa de 105 pluviòmetres.

L'efecte de la baixa densitat de pluviòmetres evidencia la diferència en la textura final d'ambdós mètodes. Així, els splines accentuen la forma arrodonida comentada anteriorment, mentre que la combinació final manté la textura característica ja esmentada.

3.2. Validació creuada

Per a validar la metodologia presentada una de les tècniques que s'ha utilitzat ha estat la validació creuada, un estàndard dins les diferents metodologies per avaluar i validar procediments.

La validació creuada consisteix a comparar, per un camp donat, el valor del pluviòmetre (referència) amb el valor que obtindríem en aquell mateix punt si apliquéssim la metodologia sense utilitzar aquell pluviòmetre. Això es fa per a tots els pluviòmetres disponibles. Com més s'assembli el registre de la referència amb el del camp reconstruït més validesa té el mètode. S'ha utilitzat com a indicador per poder comparar els resultats de la validació creuada mitjançant splines i la corresponent a la nova metodologia la correlació entre els valors de la referència i els corresponents als estimats.

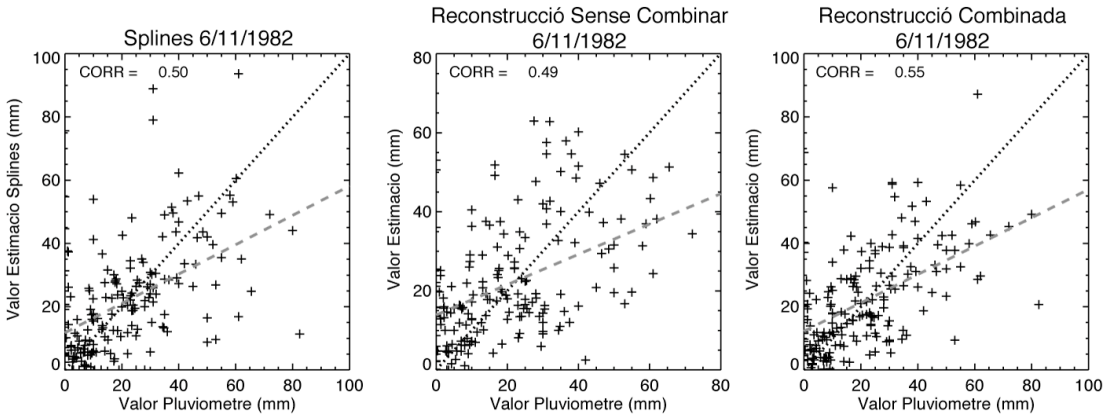

Figura 11: Diagrama de dispersió de la validació creuada per al dia 6 de Novembre de 1982. La figura de l'esquerra mostra el resultat per a l'interpolació amb splines. La figura del centre la validació amb la metodologia presentada abans d'aplicar la combinació amb pluviòmetres. La figura de la dreta mostra la validació per al camp final reconstruït amb el camp ja combinat amb els pluviòmetres. La línia puntejada representa la línia 1:1, i la línia discontinua la millor regressió lineal.

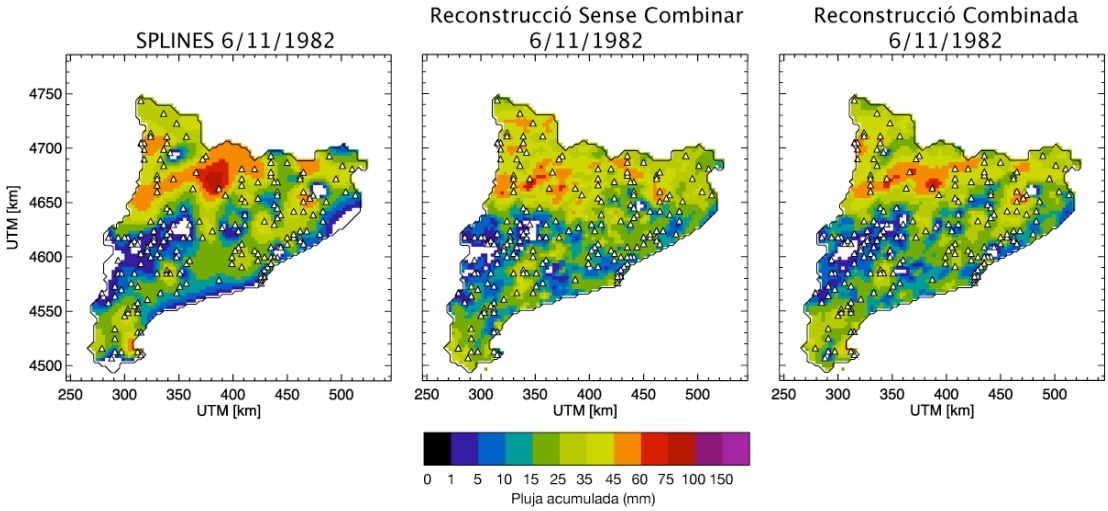

Figura 12: Ídem Figura 7 però per al dia 6 de Novembre de 1982.

En la Figura 11 es mostren els diagrames de dispersió resultants d'aplicar la validació creuada per al dia 6 de Novembre de 1982, la reconstrucció del qual es mostra en la Figura 12. Es compara per tant les mesures registrades als pluviòmetres amb els valors estimats per diferents mètodes per aquell dia. S'observa que els valors de correlació per aquest dia no són gaire elevats per cap dels mètodes presentats, però es pot veure com la correlació del mètode proposat és gairebé igual que per als splines de làmina prima abans de combinar, i supera aquesta després de la combinació (0.5 vs 0.55).

Un altre aspecte a destacar és que la reconstrucció del camp amb la nova metodologia presenta menys "outliers" (elements anormalment allunyats del núvol de punts de la regressió) que la metodologia d'splines. D'aquesta manera, les interpolació amb splines presenta certs valors que han sobreestimat molt el valor dels pluviòmetres, i a la vegada no ha estat capaç de reproduir de forma eficient el punt en que hi ha el pluviòmetre que ha enregistrat més pluja (és el punt en què el pluviòmetre va marcar 82 mm mentre que la interpolació del camp sense utilitzar aquest pluviòmetre reprodueix una pluja de 10 mm).

Finalment cal destacar que els resultats milloren de forma considerable gràcies a aplicar l'últim pas de la metodologia consistent a combinar els camps estimats amb els pluviòmetres mitjançant com s'ha comentat un interpolador exacte com és el kriging. La correlació augmenta en aquest cas de 0.49 a 0.55.

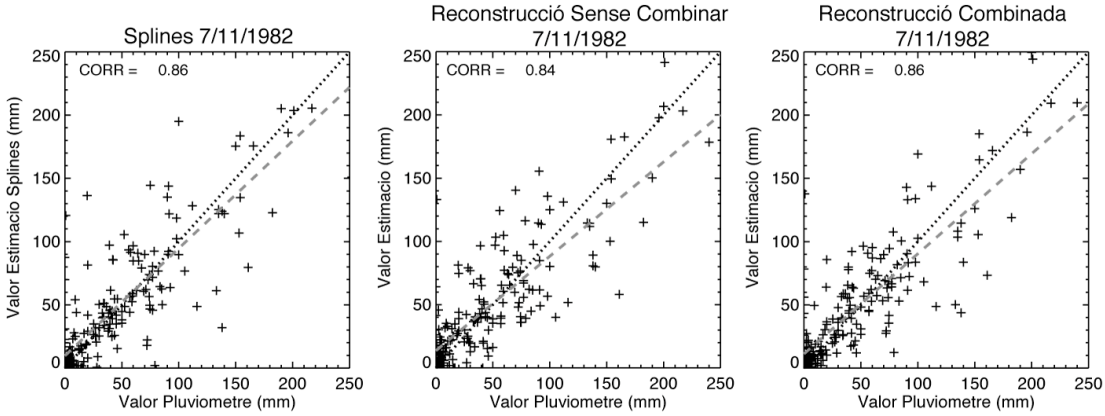

Figura 13: Ídem Figura 11 però per al dia 7 de Novembre de 1982.

La Figura 13 mostra els resultats de la validació per a l'episodi de plujes extremes del 7 de Novembre del 1982 mostrat anteriorment (Figura 8). En aquest cas les correlacions són molt més elevades que en el cas anterior, i ara ambdues metodologies tenen coeficient de correlació pràcticament idèntics. En general s'ha identificat que en els episodis amb pluja intensa els coeficients de correlació són molt elevats, i que a la vegada no hi ha grans diferències entre la nova metodologia i la interpolació clàssica amb splines.

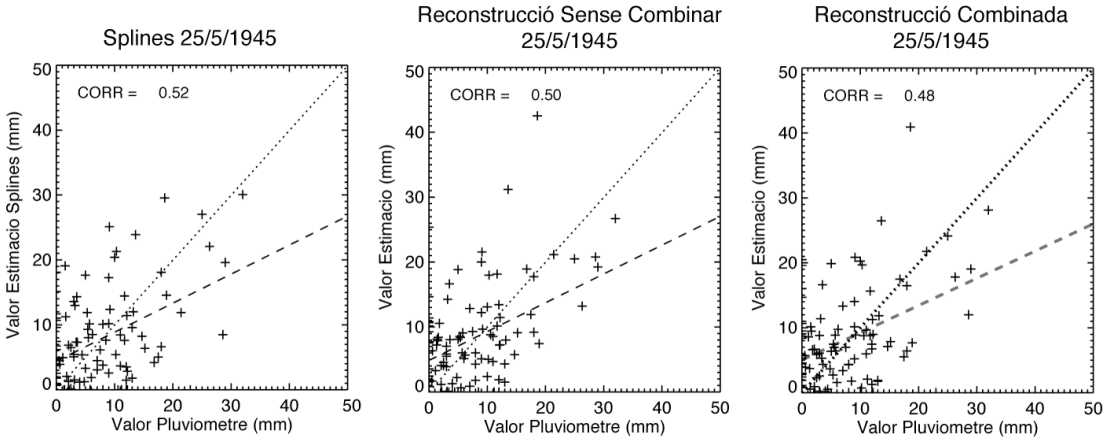

Figura 14: Ídem Figura 11 però per al dia 25 de Maig de 1945.

La Figura 14 mostra la validació d'un episodi de pluja moderada en un dia en que es disposa de pocs pluviòmetres (105). Els diagrames de dispersió que es mostren són els corresponents als camp mostrats en la Figura 10. S'observa que la correlació en aquest cas afavoreix lleugerament a la interpolació amb splines. L'aspecte dels camps, però, observats en l'esmentada Figura 10, és molt més favorable als camps reconstruïts. En general, en els dies en què es disposa de pocs pluviòmetres, s'han observat que les correlacions son baixes per a ambdues metodologies, arribant a valors extremadament baixos en casos de poca pluja. Això és degut a que la tècnica de validació creuada és molt sensible al número de

pluviòmetres disponibles, de manera quan n'hi ha pocs, una mala estimació en un sòl punt es tradueix en una ràpida disminució de la correlació en el conjunt.

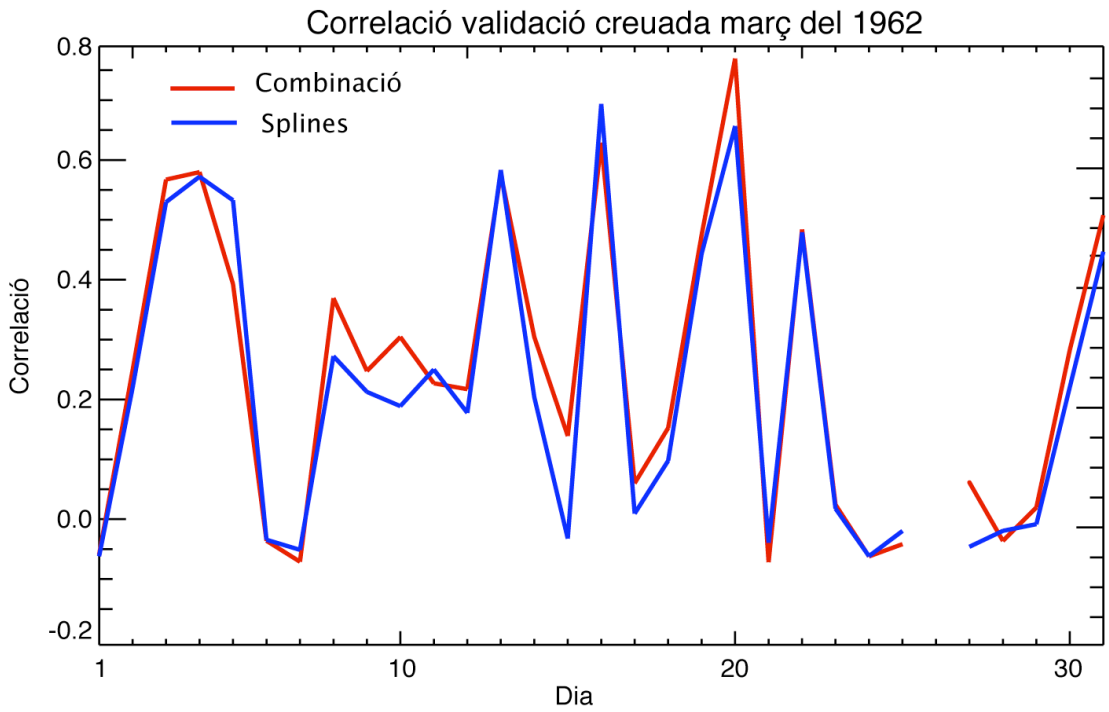

Figura 15: Coeficients de correlació obtingut aplicant la metodologia de validació creuada a tots els dies del mes de març del 1962. En blau es mostra la correlació corresponent a comparar la referència amb l'obtinguda mitjançant la metodologia proposada i en blau en comparar-la amb l'obtinguda amb els splines.

Per avaluar si realment hi ha diferències significatives s'ha aplicat la metodologia de la validació creuada per a tots els dies de l'any del 1962. Els resultats en efecte no mostren excessives diferències entre els coeficients de correlació obtinguts de comparar la referència amb la reconstrucció mitjançant el mètode proposat i la correlació obtinguda en comparar-la mitjançant splines, essent aquesta a vegades superior per una metodologia i a vegades per a l'altre. De fet, la mitjana de les correlacions al llarg de l'any son pràcticament idèntiques (0.22 per a la nova metodologia vs 0.23 per a les splines). La Figura 15 presenta una mostra de l'any corresponent al mes de Març. En ella s'observa que, si bé la correlació sembla ser més cops favorable a la comparativa utilitzant els camps combinats, també hi ha casos que és a l'inrevés, de manera que no es pot concloure la correlació sigui de forma general més favorable a una o a l'altre . A més no s'observat cap patró a partir del qual es pugui decidir en quins casos s'afavoreix cada metodologia, el que també s'evidencia en l'estudi per a tot l'any. Els dies en què no hi ha valor en la gràfica és perquè els algun dels camps té valor zero i per tant no se'n pot calcular la correlació

D'aquesta manera es pot argumentar que qualitativament els camps reconstruïts sempre tenen millor aspecte que els interpolats amb splines, i que això s'evidencia quan hi ha pocs pluviòmetres. En canvi, si quantifiquem això en termes de correlació en un anàlisi de validació creuada, els resultats són molt semblants, i les lleugeres diferències a favor d'un o altre s'han prendre amb cautela.

3.3. Comparació amb referència externa

Una altra metodologia emprada per a validar la metodologia exposada ha estat la de comparar els camps reconstruïts amb aquesta metodologia amb una referència externa. Com a referència externa s'ha usat els camps obtinguts mitjançant la combinació de les estimacions de pluja del radar meteorològic de l'Agència Estatal de Meteorologia (AEMET) amb els valors puntuals dels pluviòmetres mitjançant al tècnica ja comentada de kriging amb deriva externa (Velasco-Forero et al. 2009). Com s'ha exposat en l'apartat 2.1.2, considerem els camps resultants com la millor referència possible.

Cal aclarir en aquest punt que només es disposa dels valors del radar de l'AEMET per a episodis molt concrets de pluja durant pocs dies dels anys 2001 al 2005. Per tant la metodologia exposada segueix sent necessària per a tot el període proposat.

D'aquesta manera es disposa d'una sèrie d'esdeveniments de pluja on avaluar la nova tècnica i comparar els resultats en front de la ja comentada metodologia d'interpolació dels valors puntuals dels pluviòmetres per splines de làmina prima.

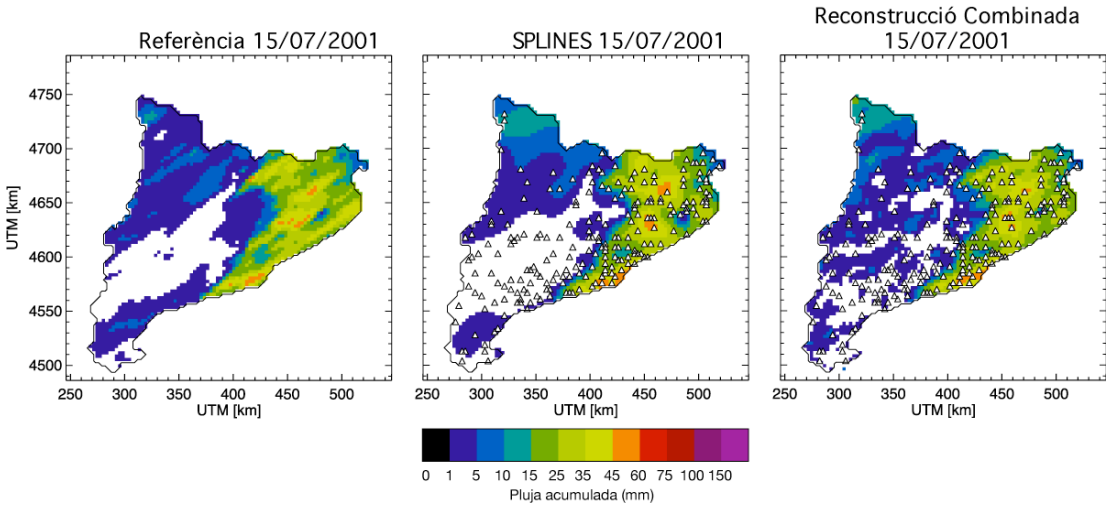

Figura 16: Camps de pluja per al dia 15 de Juliol de 2001. A l'esquerra es pot veure el camp de referència (radar AEMET més combinació amb pluviòmetres), al centre el camp obtingut mitjançant la interpolació de pluviòmetres amb splines, mentre que a la dreta es pot observar la reconstrucció mitjançant la tècnica presentada.

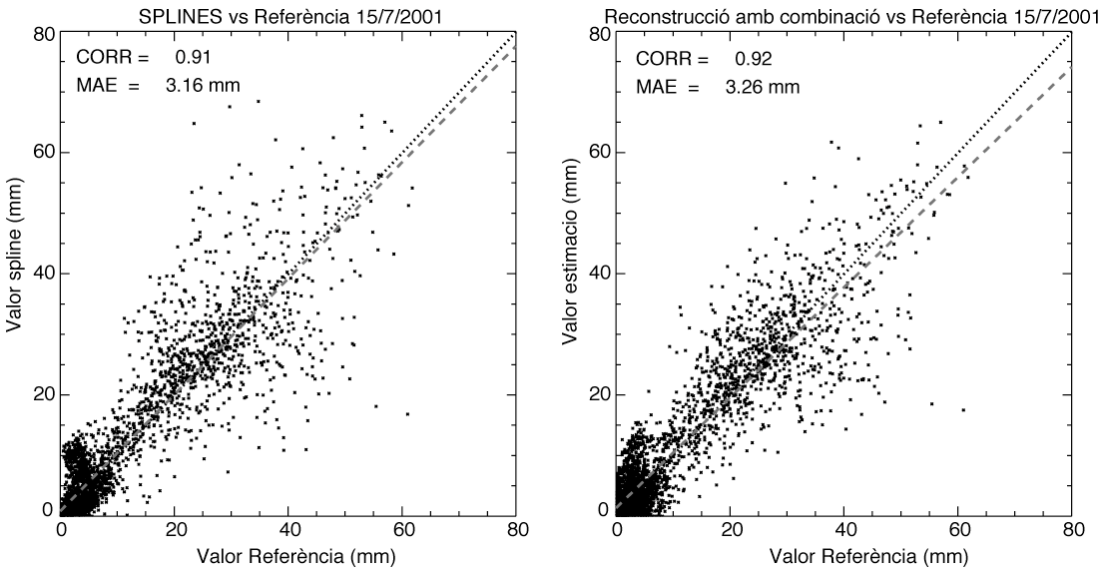

Figura 17: Diagrames de dispersió per al dia 15 de Juliol de 2001. A l'esquerra es mostra el gràfic corresponent a comparar el camp obtingut mitjançant interpolació per splines amb el camp de referència; de forma similar, a la dreta es mostra el gràfic corresponent al camp obtingut amb la nova metodologia. La línia puntejada representa la línia 1:1, i la discontínua la millor regressió lineal. La correlació i l'error absolut mitjà (MAE: Mean Absolut Error) de cada cas estan escrits al gràfic corresponent.

La Figura 16 mostra els diferents camps de pluja per al dia 15 de Juliol de 2001. Per aquest dies es disposa d'un xarxa densa de pluviòmetres, concretament per aquest dia s'han utilitzat 221 pluviòmetres. Com s'ha comentat ja anteriorment, les splines mostren unes formes més arrodonides, mentre que la reconstrucció amb la nova metodologia aconsegueix reproduir una textura semblant a la de la referència. De nou, però, com ja s'ha vist en altres casos (veure Figura 9) la reconstrucció indica pluja en localitzacions on la referència clarament no ho fa (comarques de la Segarra, Pla d'Urgell). Val a dir, però, que es tracta de quantitats minses de l'ordre de 1mm que quantitativament no son importants per bé que visualment no son desitjables.

L'anàlisi comparatiu el trobem a la Figura 17. En ella s'ha representat el diagrama de dispersió del valor de referència (radar de AEMET + combinació) en front de cadascuna de les reproduccions mitjançant les dues metodologies comparades: la interpolació per splines i la tècnica exposada en aquest projecte. A més, per a poder-los comparar de forma numèrica, s'han considerat dos paràmetres: la correlació i el MAE o error absolut mitjà. El MAE es defineix com la mitjana del valor absolut de les diferències entre tots els valors observats respecte als estimats, i per tant indicarà un millor comportament com més petit sigui.

En els gràfics presentats observem unes correlacions molt altes i pràcticament iguals en ambdues metodologies (0.91 en splines vs 0.92 en reconstrucció). El MAE és relativament petit i novament molt similar per a les dues tècniques.

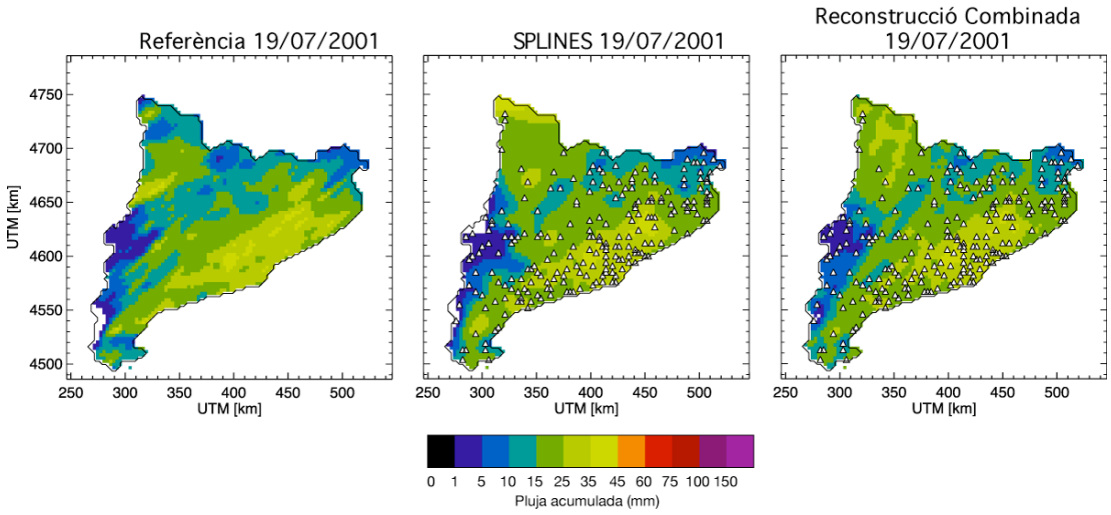

Figura 18: Ídem Figura 16 però per al dia 19 de Juliol de 2001.

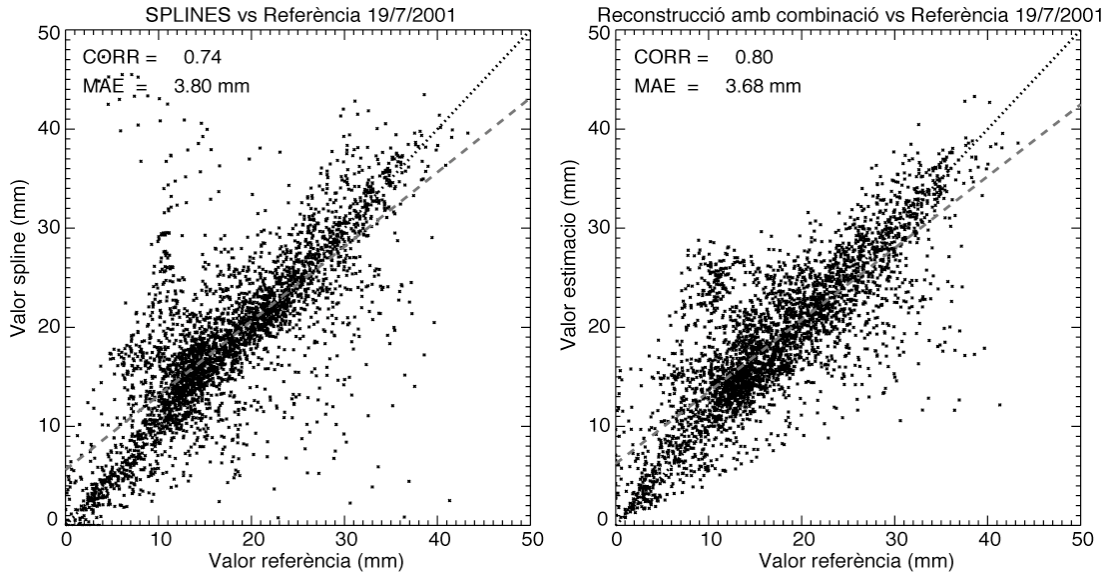

Figura 19: Ídem Figura 17 però per al dia 19 de Juliol de 2001.

En la Figura 18 es mostren els camps per al dia 19 de Juliol del 2001, mentre que els seus diagrames de dispersió els trobem a la Figura 19. Les correlacions en aquest cas no són tant altes, essent lleugerament superior la corresponent a la reconstrucció mitjançant la nova metodologia (0.74 vs 0.80). El MAE afavoreix també de forma lleugera a la reconstrucció amb combinació (3.80 mm vs 3.68 mm).

El més interessant, però, és veure com les splines presenten punts molt allunyats de la referència, els ja comentats "outliers", deguts a alguna zona de divergència d'aquest interpolador. Trobem un número considerable de punts en què la referència indica valors d'entre 5 i 15 mm i l'estimació de l'spline entre 40 i 50 mm. A la vegada hi ha alguns punts força plujosos que l'spline ha subestimat. Aquest efecte és molt menys acusat en el cas de la reconstrucció mitjançant la nova metodologia presentada.

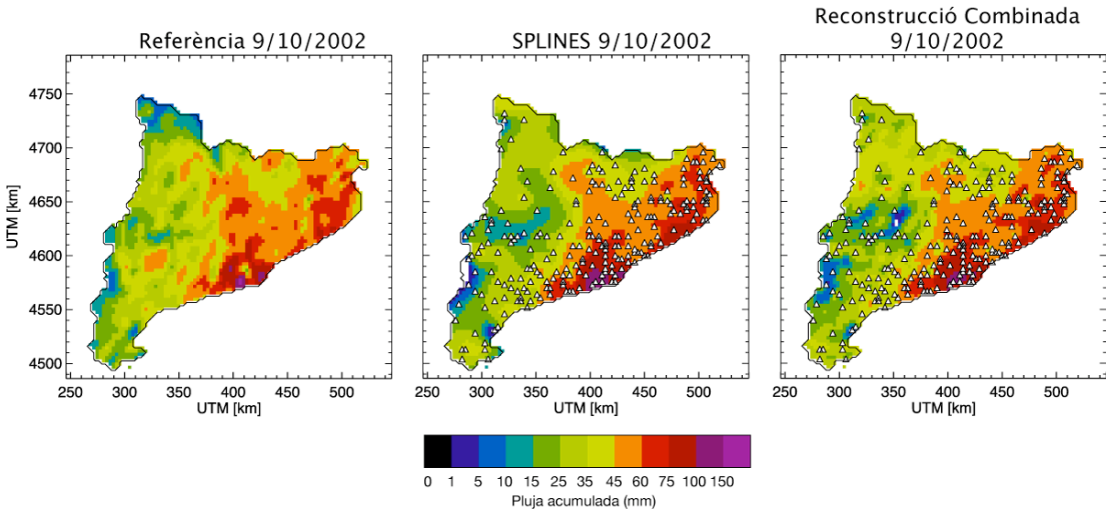

Figura 20: Ídem Figura 16 però per al dia 9 d'Octubre del 2002.

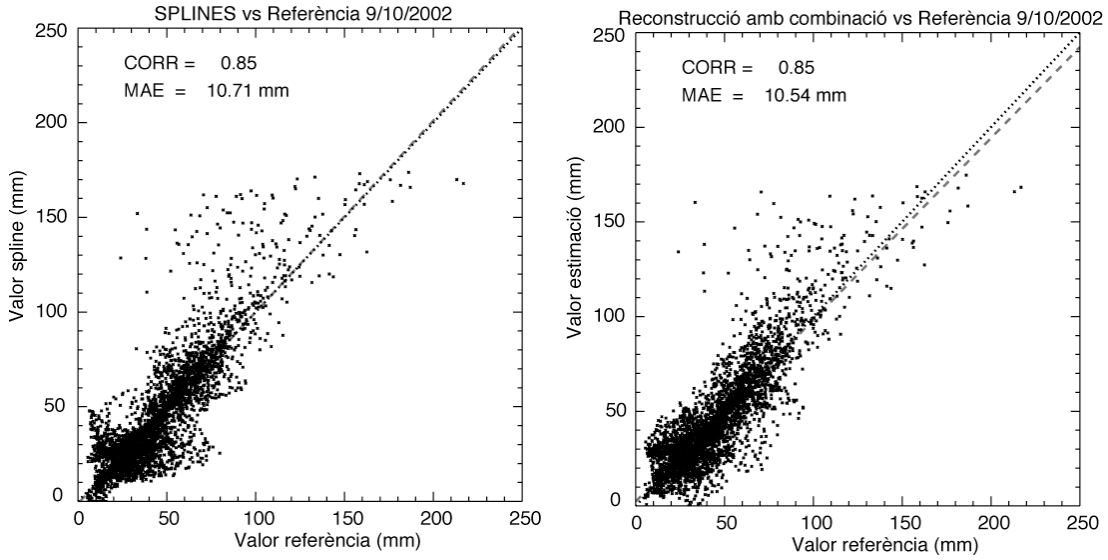

Figura 21: Ídem Figura 17 però per al dia 9 d'Octubre del 2002.

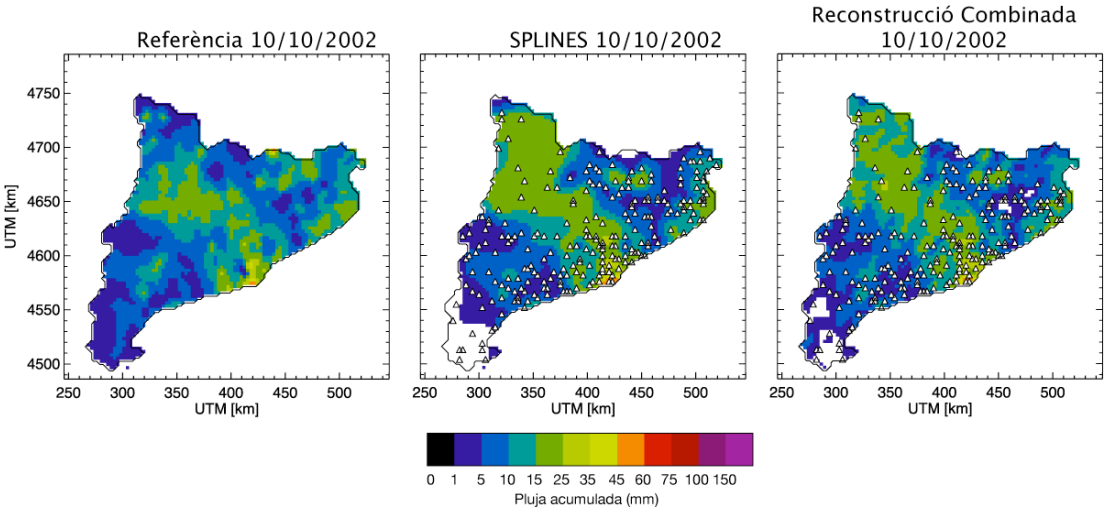

Figura 22: Ídem Figura 16 però per al dia 10 d'Octubre del 2002.

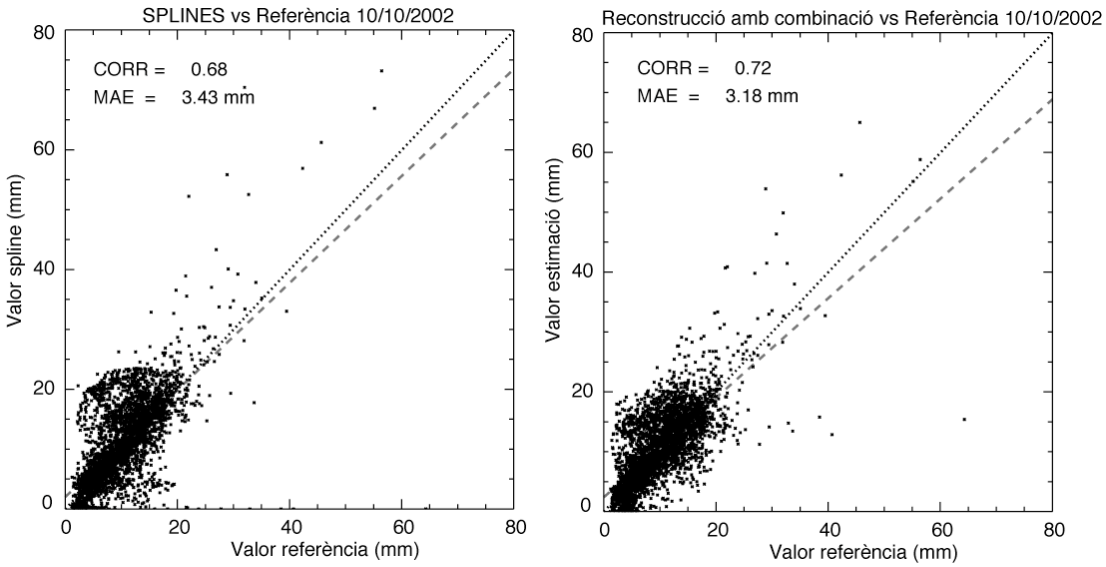

Figura 23: Ídem Figura 17 però per al dia 10 d'Octubre del 2002.

S'ha realitzat també la comparativa per a l'episodi de fortes plujes al Baix Llobregat i Barcelonès de l'Octubre de 2002 (Figura 20 i Figura 22 per als camps de pluja; Figura 21 i Figura 23 per als gràfics de dispersió) Per al dia 9 d'Octubre, un episodi amb acumulacions extremes de més de 200 mm, els valors de correlació i MAE són gairebé idèntics. Val a dir que en els episodis amb valors més alts de pluja el MAE és lògicament major , ja que aquest té en compte les diferències entre estimació i observació de forma absoluta i no relativa. L'episodi del 10 d'Octubre per la seva banda presenta un coeficient de correlació moderadament superior per a la tècnica presentada i un MAE també menor.

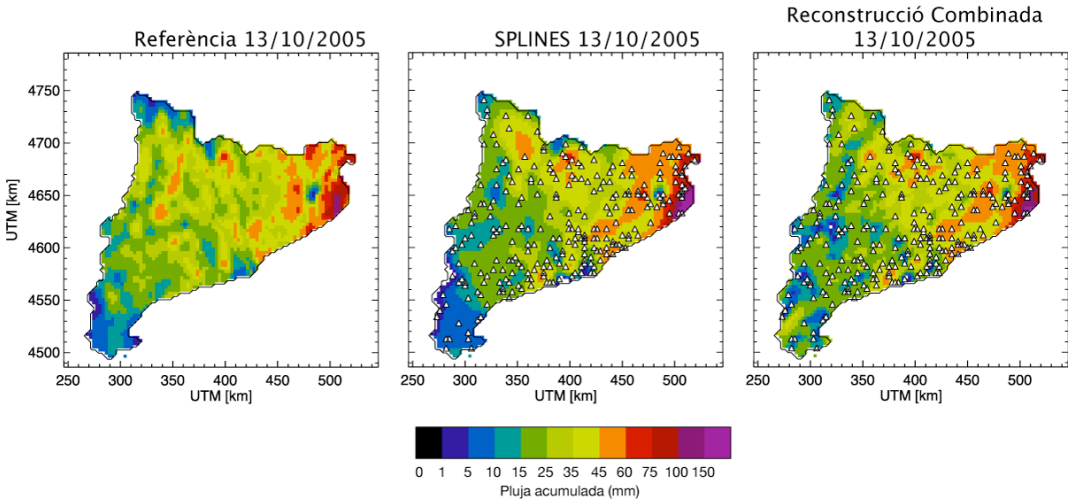

Figura 24: Ídem Figura 16 però per al dia 13 d'Octubre del 2005.

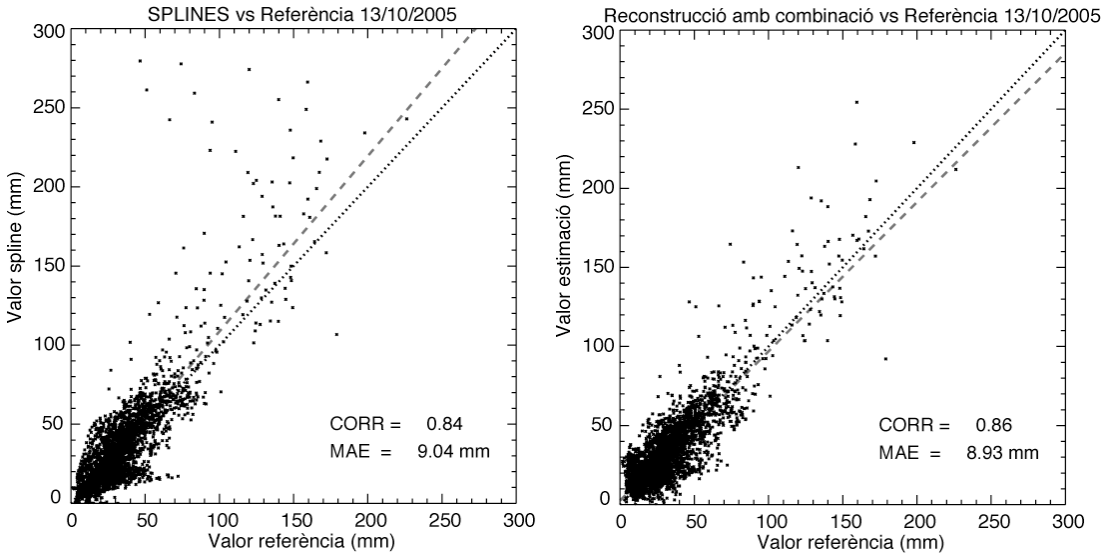

Figura 25: Ídem Figura 17 però per al dia 13 d'Octubre del 2005.

La Figura 24 presenta el cas de l'episodi que va provocar inundacions al Baix Empordà l'Octubre del 2005. La Figura 25 correspon als gràfics de dispersió per aquest dia, on novament s'observen coeficients de correlació i MAE lleugerament millors per a la nova metodologia. Novament cal destacar, com hem comentat per als gràfics de dispersió del dia 19 de Juliol de 2001 (Figura 19), un elevat número de punts allunyats de la recta 1:1, en aquest cas degut a que la interpolació mitjançant splines ha sobreestimat els valors de la referència. Això és degut a la forma com les splines extenen els valors extrems de pluja. Aquest efecte no es produeix en la nova metodologia, el que fa que hi hagi menys outliers.

4. Quantificació del temps emprat

El present treball s'emmarca dins els estudis de Doctorat en Enginyeria Civil que desenvolupa l'autor.

El projecte s'ha realitzat durant el període de 7 mesos que va des del Gener del 2010 al Juliol del 2010 amb una dedicació exclusiva.

5. Conclusions

S'ha desenvolupat una metodologia que treu profit dels avenços tecnològics més actuals quant a estimació de pluja per tal de millorar les reconstruccions en el passat. Aquesta metodologia s'ha aplicat de forma satisfactòria per a tot el període d'estudi que va des del 1941 al 2007.

Els primers resultats d'aplicar la tècnica presentada per a reconstruir els camps de pluja diaris des del 1941 mostren camps de pluja amb una variabilitat espacial superior a mètodes tradicionals com els splines de làmina prima. Aquest fet s'observa en una "textura" del camp de precipitació molt més real, eliminats els clàssics efectes de "bombolla" i gradients erronis (divergències) del mètode d'splines.

S'ha realitzat un anàlisi mitjançant la metodologia de validació creuada per a episodis específics i per a un any sencer de dades, el 1962. La metodologia conclou de forma general que, si bé en general el mètode proposat obté resultats lleugerament superiors a l'obtingut mitjançant la interpolació per splines, aquests no són en cap cas definitoris. D'aquesta manera en l'anàlisi per a l'any 1962 s'observa que l'indicador utilitzat (coeficient de correlació entre valor real i estimació) és a vegades més favorable a la nova metodologia i a vegades als splines, sense que s'hagi pogut observar un patró de comportament que conclogui quan un és superior a l'altre.

Com a segona metodologia de verificació de la tècnica s'ha realitzat la comparació entre els resultats obtinguts mitjançant la tècnica descrita i un camp de referència externa, en aquest cas el radar de l'AEMET de Corbera de Llobregat per als episodis disponibles.

Els resultats obtinguts han estat satisfactoris. La textura del radar es preserva i els camps son visualment molt semblants als observats mentre que els coeficients de correlació entre observació i estimació son força alts. La comparativa d'aquests indicadors amb els que s'obtenen si s'utilitza com a camp estimador la metodologia comentada dels splines afavoreix en tots els casos estudiats la nova metodologia. Adicionalment, la tècnica proposada presenta menys punts allunyats de forma clara del núvol de la regressió (els anomenats "outliers"), el que si s'observa de forma clara en els diagrames de dispersió obtinguts mitjançant splines.

Finalment cal destacar el diferent recorregut que han de tenir els resultats obtinguts mitjançant una metodologia clàssica com la de les splines, utilitzada en tot l'estudi com a referència on comparar-se, i la nova tècnica mostrada. Mentre que la metodologia clàssica està del tot desenvolupada, la reconstrucció obtinguda mitjançant la metodologia presentada té encara un llarg recorregut de millora. Tant

a nivell de millora en la metodologia pròpiament dita, fent ús de noves tècniques i millorant l'eficiència de les que s'usen (veure apartat següent, 5.1) com per l'increment del volum de dades per a calibrar el sistema que es produirà, tant de forma natural a mesura que avanci el temps i es tinguin més mesures de camps de radar com per la utilització de la totalitat de les dades radars existents però no plenament disponibles.

Com a conclusió final podem dir que la metodologia presentada millora els mètodes presents fins ara per a l'obtenció de camps històrics de pluja a Catalunya, i que la seva aplicació ha permès obtenir una sèrie històrica de camps de pluja que resulten molt útils per a totes les àrees d'estudi que necessitin camps diaris de precipitació acumulada d'alta resolució.

5.1. Perspectives

Hi ha diversos aspectes a explorar que sens dubte ajudaran a millorar la reconstrucció dels camps diaris.

El primer és realitzar una anàlisi encara més exhaustiva de les dades disponibles, sobretot de la sèrie de dades històrica de pluviometria. La informació històrica disponible és la base sobre la qual gira qualsevol metodologia de reconstrucció, de manera que l'anàlisi a nivell diària de la validesa i qualitat dels registres dels pluviòmetres contribuirà sens dubte a obtenir camps més fidedignes.

La metodologia també presenta punts a investigar per tal de millorar-la. Un sens dubte és la regressió multivariant. Cal explorar tècniques que ajudin a eliminar els problemes de colinealitat esmentats en l'apartat 2.3.2, com ara la de la regressió pas a pas (*Stepwise regression*, (Storch; Zwiers 1999) o la regressió esbiaixada (*Ridge regression* (Birkes; Dodge 1993). A la vegada seria interessant fer un anàlisi de sensibilitat i establir una metodologia que permeti escollir de forma dinàmica per a cada dia el número de Components Principals que s'utilitzen en la regressió en funció del número de pluviòmetres disponibles (actualment agafem un número de constant de CP).

Un altre aspecte a estudiar té a veure amb l'obtenció dels patrons de pluja durant la calibració. Per aquest estudi s'han extret un únic conjunt de patrons per a tot l'any (resultants de l'Anàlisi de Components Principals), però es podria pensar en diferenciar patrons per a dies amb pluja generalitzada de dies amb poca pluja, o bé definir diferents patrons en funció de la tipologia de precipitació, diferenciant per tant els casos de pluja de tipus convectiva dels casos de pluja de tipus estratiforme, o fins i tot diferenciar patrons en funció de l'estació de l'any.

Finalment cal destacar que el període de calibració amb el que s'ha treballat ha estat aquell en què les dades radar eren disponibles i tenien suficient qualitat. Si es disposés de la totalitat de les dades dels diferents radars durant el temps en què

aquests han estat operatius el període de calibració es podria ampliar dels 2 anys actuals a 7 o 8, en funció de la validesa de les dades. Això incrementaria de forma dràstica la robustesa de l'obtenció de patrons, i ajudaria a la vegada a realitzar l'esmentada diferenciació d'aquests en subconjunts.

6. Referències

Ansell, T. J., 2006: Daily Mean Sea Level Pressure Reconstructions for the European North Atlantic Region for the Period 1850-2003. *Journal of Climate*, **19**.

Berenguer, M., 2006: Validation of algorithms for rainfall estimation and forecasting. Hydrological uses of meteorological radar, Universitat Politècnica de Catalunya, 130 pp.

Birkes, D., i Y. Dodge, 1993: *Alternative Methods of regression.* Wiley-Interscience, 228 pp.

Deutsch, C. V., i A. G. Journel, 1997: *GSLIB: Geostatistical Software Library and User's Guide.* Second Edition ed. Oxford University Press, 368 pp.

Franco, M., 2008: Corrección del error asociado a la variación vertical de la reflectividad. Estimación cuantitativa de la lluvia mediante radar meteorológico, Universitat Politècnica de Catalunya, 254 pp.

Isaaks, E. H., i R. M. Srivastava, 1989: *Applied Geostatistics.* Oxford University Press.

Kaplan, A., A. Kushnir, M. A. Cane, i M. B. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysycal Research*, **102**.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. *Science Report 1, Statistical Forecasting Project, Department of Meteorology*, MIT, 49.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, i A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, **108**.

Rosenfeld, D., D. Atlas, D. B. Wolff, i E. Amitai, 1994: The window probability matching method for rainfall measurements with radar. *Journal of Applied Meteorology*, **34,** 682-693.

Sánchez-Diezma, R., D. Sempere-Torres, G. Delrieu, i I. Zawadzki, 2001: An Improved Methodology for ground clutter substitution based on a pre-classification of precipitation types. *30th International Conference on Radar Meteorology*, Munich, Germany, 271-273.

Schiemann, R., C. Frei, i M. A. Liniger, 2010: Reduced-space optimal interpolation of daily rain-gauge precipitation in Switzerland. *Journal of Geophysical Research*, **(acceptat)**.

Schumaker, L., 2007: *Spline Functions: Basic Theory.* Cambridge University Press.

Searcy, J. K., i C. H. Hardison, 1960: Double-mass curves. *U.S. Geological Survey Water-Supply*, **1541-B**.

Sempere-Torres, D., J. M. Porrà, i J. D. Creutin, 1994: A general formulation for rain drop size distribution. *J. Appl. Meteor.*, **33,** 1494-1502.

Storch, H. V., i F. W. Zwiers, 1999: *Statistical analysys in climate research.* Cambridge University Press, 484 pp.

Velasco-Forero, C. A., D. Sempere-Torres, E. F. Cassiraga, i J. J. Gómez-Hernández, 2009: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Advances in Water Resources*, **7**.

Wilks, D. S., 2006: *Statistical methods in the atmospheric sciences.* 2 ed. Academic Press.

Wilson, E. M., 1990: *Engineering Hydrology.* 4 ed. Macmillan Press, 348 pp.

Wilson, J. W., i E. A. Brandes, 1979: Radar measurement of rainfall- a summary. *Bull. Amer. Meteor. Soc.*, **60,** 1048-1058.